PRIOR INFORMATION AND ILL-CONDITION IN REGRESSION SYSTEMS

by

E. L. Battiste and R. L. Anderson

Institute of Statistics
Mimeo Series No. 534
Spring 1967

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

1. INTRODUCTION

1.1. Statement of the Problem

Engineers, faced with the problem of building mathematical models of physical processes, utilize past experience in addition to current experiments for aid in model specification and parameter estimation. When such past experience is available, its use in estimation conditions the estimation equations in a positive manner. Engineers realize that such augmentation of the model will bias estimates, while hopefully decreasing their mean square error, if the knowledge added is incorrect. The effect that this prior knowledge exerts on properties of estimates, properties of the resultant predictive model, and on power of tests of hypotheses, should be investigated.

This dissertation examines the least squares analysis technique in general. It is assumed that the experimenter has, in addition to his current experimental results, modeled in the form

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}, \quad \underline{\varepsilon} \sim MN(\underline{0}, \sigma^2 I_n), \tag{1.1}$$

other information available on $\underline{\beta}$, termed

$$\underline{\beta}_0 = \underline{\beta} + \underline{\delta}\ , \tag{1.2}$$

where $\underline{\delta}$ is distributed with mean $\underline{0}$ and variance $D\sigma^2$. That is, the variance-covariance matrix of $\underline{\delta}$ is $D\sigma^2$.

However, there is another situation in which prior knowledge is used to which the discussions given here apply equally well. Consider a given experiment, where the model is considered to be of the form (1.1), where $\underline{\varepsilon}$ reflects measurement, model, and experimental errors.

Historically, we observe that $\underline{\beta}$ tends to vary about a prior $\underline{\beta}_0$ as follows:

$$(\underline{\beta}_0 - \underline{\beta}) = \underline{\delta} \qquad (1.3)$$

where $\underline{\delta}$ reflects this variation. That is, $\underline{\beta}_0$ is another estimator of $\underline{\beta}$ with variation reflected by $\underline{\delta}$. $\underline{\delta}$ may have mean $\underline{0}$ or mean $\underline{m}$, the latter reflecting bias in the prior. The variance of $\underline{\delta}$ is a measure of the usefulness and validity of $\underline{\beta}_0$.

The models and estimation schemes in this dissertation apply to both of the above situations. Regression procedures with added prior information of various types are studied.

Questions considered are:

1. What effects on mean square error, variance, and bias of estimators do various types of prior knowledge exert, if that knowledge is correct?

2. How robust are the properties of the estimators to the utilization of incorrect prior knowledge, and how is this robustness dependent on correlations between the independent variables?

3. What changes in the power of tests of hypotheses can be expected?

When prior information is not available, a technique due to A. E. Hoerl[1], of E. I. dePont de Nemours and Company, is valuable if the regression model is ill-conditioned. This technique is analogous to adding prior knowledge of a certain type to the regression model.

---

[1]This technique is described in an unpublished monograph by A. E. Hoerl entitled "Ridge Regression".

Under the conditions required in that analogy, the resulting changes in power of tests of hypotheses are examined.

If the prior knowledge can be assumed to be normal in distributional form, maximum likelihood estimators give the same results, for the fixed parameter model, as do proper applications of the Gauss-Markov Theorem; the resultant estimators are best linear unbiased estimators. This would not be the case if other prior distributions were applied.

When deciding on the appropriateness of the use of prior information, any decision is based on the validity of the prior knowledge. The equations which state the decision procedure in terms of mean square error are generally difficult to use. This difficulty is met by viewing varied situations using a computer program. Evaluation of exact power of tests of hypotheses leads to a problem for which only approximate solutions exist. In this thesis, this difficulty is met by simulating the true situation.

1.2. Literature Review

Many papers have discussed topics allied to the discussion of prior knowledge as given in this paper. Among those which take a non-Bayesian view of the subject are the papers by Durbin (1953), Wells and Anderson (1964)[2], Chipman (1963), and Dykstra (1966). After

---

[2]Also presented at the Eastern Regional meeting of The Institute of Mathematical Statistics and American Statistical Association, Florida State University, Tallahassee, Florida, April, 1965. Regression Procedures for Missile Trajectory Estimation, By W. T. Wells, Wolf Research and Development Corporation, and R. L. Anderson, North Carolina State University at Raleigh.

development of ridge analysis techniques [Hoerl (1959), and Draper (1963)], Hoerl (1962) worked on their application to the ill-conditioned regression problem.

Durbin applied extraneous information to a model in the manner followed in this paper, acquiring estimators with the properties available from least squares procedures, but with lower variance.

Wells and Anderson considered maximum likelihood estimation for a model with random mean, where prior knowledge was available in the form of the average and variance of this mean. In this case, statements could be made about the relationship between the incorrectness of this prior knowledge and its value to the experimenter.

Chipman discussed minimum mean square error estimation, and showed when this method reduces to the method of least squares. He imposed incorrect estimable linear restrictions on the model. These gave estimators which, when biased, had smaller variances.

Dykstra proposed methods for specifying additional runs to be added to the data of a non-orthogonal experiment, in such a way as to decrease the statistical dependence among the controllable variables.

All of these techniques condition the regression model in a manner which reduces variances of parameter estimates. To date no investigation has been carried out pertaining to the addition of incorrect prior knowledge, its relationship to the condition of the model, to mean square error, to variance estimates, to predictive ability of the model, and to power of tests of hypotheses.

The practice of adding prior knowledge is well established. The study of this practice is needed, and augmentation of the study given in this paper is called for.

2. THE REGRESSION MODEL WHEN PRIOR INFORMATION IS UTILIZED

2.1. Introduction and Notation

In this chapter, models are examined mathematically under the condition that prior information is available about the parameters. This prior information is considered to be normal in form, but information in terms of mean and variance alone would, with the use of the Gauss-Markov Theorem, give the same results. Section 7.1 discusses the development of this prior knowledge from previous regressions.

Predictive ability and properties of estimators are considered, as are properties of estimators when the prior information is incorrect. Computations performed, and described in Chapter 4, show how properties relate to different correlations between the independent variables. Simulations performed relate conditioning, or the addition of prior knowledge, to the power of tests of hypotheses. A computer program compares estimators resulting from the use of prior knowledge with those obtained by not using prior knowledge.

In this dissertation it is assumed that the experimenter bases his estimates on the following model[3]:

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}, \quad \underline{\varepsilon} \sim MN(\underline{0};\ \sigma^2 I_n);$$
$$\underline{\beta}_0 = \underline{\beta} + \underline{\delta}, \quad \underline{\delta} \sim MN(\underline{0};\ \sigma^2 D); \qquad (2.1)$$

[3]In general one has prior information on q of the parameters $(q \leq p)$. We will consider only the case q=p in this dissertation.

where D is of order p, $\underline{y}$ has n elements, $\underline{\beta}_0$ has p elements, and X is n by p. The second equation states the assumed prior knowledge about $\underline{\beta}$. If prior knowledge is also available on σ^2, the following equation is added to the model:

$$c/\sigma^2 \sim \chi^2_k \quad , \tag{2.2}$$

where χ^2_k is the χ^2 variate with k degrees of freedom. Added knowledge of any kind is always to be considered as independent information. That is, $E(\varepsilon_i\delta_j) = E(\varepsilon_i)E(\delta_j)$ for every i,j. Independence also holds for information on σ^2.

Estimators and their properties will use the notation (^) when no prior knowledge is used and (~) when prior knowledge (2.1) or (2.1) and (2.2) is applied.

The distributions from which the prior knowledge actually comes are stated as follows. Either

a. $\underline{\beta}_0 = \underline{\beta} + \underline{m} + \underline{\delta}$, where $\underline{\delta} \sim MN(\underline{0};\ \sigma^2 D_0)$, so that

$$g_{01}(\underline{\delta}|\sigma^2) = \frac{|D_0|^{-1/2}}{(2\pi\sigma^2)^{p/2}} \exp\left[\frac{-1}{2\sigma^2} \quad \underline{\delta}' D_0^{-1} \underline{\delta}\right] , \tag{2.3}$$

b. $\beta_{0j} = \beta_j + m_j + \delta_j$, where δ_j is distributed as a beta variate with zero mean and variance d_{0j}, in which case

$$g_{02}(\delta_j|\sigma^2) = \frac{\Gamma(8)\,[3\sqrt{d_{0j}}-\delta_j)(3\sqrt{d_{0j}} + \delta_j)]^3}{[\Gamma(4)]^2(6\sqrt{d_{0j}})^7} , \quad \delta_j \varepsilon (-3\sqrt{d_{0j}}, 3\sqrt{d_{0j}}) \tag{2.4}$$

c. β_{0j} is as in a and b above, and δ_j is distributed uniformly as in the beta case, so that

$$g_{03}(\delta_j|\sigma^2) = (2\sqrt{3d_{0j}})^{-1}, \quad \delta_j \varepsilon (-\sqrt{3d_{0j}},\ \sqrt{3d_{0j}}) , \tag{2.5}$$

or d. $\beta_{0j} = \beta_j + m_j + \delta_j$, and δ_j is an exponential deviate with zero mean and variance d_{0j}. Then

$$g_{04}(\delta_j|\sigma^2) = \frac{1}{\sqrt{d_{0j}}} \exp(-1-\delta_j/\sqrt{d_{0j}}); \; d_{0j} > 0; \; \delta_j \geq -\sqrt{d_{0j}} . \tag{2.6}$$

When prior information is available on σ^2, as in (2.2) above, distributions for $\underline{\delta}$ are given as in (2.3) through (2.6) and either

e. $c/\sigma^2 \sim \chi^2_{k_0}$ so that

$$h_{01}(c|\sigma^2) = \frac{c^{(k_0/2)-1} e^{-c/(2\sigma^2)}}{2^{k_0/2} \Gamma(\frac{k_0}{2}) \sigma^{k_0}} \quad ; \quad c > 0 \tag{2.7}$$

or f. c/σ^2 is distributed exponentially with mean k_0, in which case

$$h_{02}(c|\sigma^2) = \exp[-c/(k_0\sigma^2)]/(k_0\sigma^2) \; ; \quad k_0 > 0; \; c \geq 0 . \tag{2.8}$$

The development of the above prior densities is discussed in Appendix 7.1.

For the estimators actually used by the experimenter, with prior knowledge possibly in error, properties attained when the above priors hold are annotated with a subscript w, e.g., $E_w\tilde{\underline{\beta}}$.

Comparisons require that the following be developed:

(1) Estimation not using priors: $\hat{\underline{\beta}}$; $\hat{\sigma}^2$; $E(\hat{\underline{\beta}})$; $E(\hat{\sigma}^2)$; $V(\hat{\underline{\beta}})$, the variance-covariance matrix of $\hat{\underline{\beta}}$; $V(\hat{\sigma}^2)$; $MSE(\hat{\underline{\beta}})$, the mean square error matrix for $\hat{\underline{\beta}}$; $MSE(\hat{\sigma}^2)$; and $V(\hat{\underline{y}}_0)$; the variance-covariance matrix of the predicted values.

(2) Estimation using assumed priors when assumed priors are correct: Replace ^ by ~ in (1).

(3) Properties actually attained by the experimenter when the priors used in (2) are incorrect, i.e., the correct priors are those described in (2.3) through (2.8): E is replaced by E_W in (2), for example.

(4) When prior knowledge on σ^2 is used in estimation, $E(\tilde{\sigma}_c^2)$, $V(\tilde{\sigma}_c^2)$, and $MSE(\tilde{\sigma}_c^2)$ are the notations used.

For simplicity and clarity in expressions, especially in writing complicated properties of estimates, notational usage will be

$$S = (X'X),$$

$$\tilde{S} = (X'X + D^{-1}) = (X^{*\prime}X^{*}).$$

The models actually used by the experimenter include the following:

(1) For normally distributed current information,

$$f(\underline{\varepsilon}|\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp[-\underline{e}'\underline{\varepsilon} \;/\; (2\sigma^2)] \quad . \tag{2.11}$$

(2) For added normally distributed prior information,

$$g_1(\underline{\delta}|\sigma^2) = \frac{|D|^{-1/2}}{(2\pi\sigma^2)^{p/2}} \exp[-\underline{\delta}'D^{-1}\underline{\delta} \;/\; (2\sigma^2)] \quad . \tag{2.12}$$

(3) For added information about σ^2,

$$h_1(c|\sigma^2) = \frac{c^{\frac{k}{2}-1} e^{-\frac{c}{2\sigma^2}}}{2^{\frac{k}{2}}\,\Gamma(\frac{k}{2})\,\sigma^k} \quad ; \qquad c > 0 \quad . \tag{2.13}$$

Thus, the situation described above can be restated as follows. An experimenter has prior information about the parameters of a linear regression model. He uses this prior information, considering that it is normally distributed and that the model is (2.1), or (2.1) and (2.2). Estimates and their properties are available to him. However, if the added information is not correct, these properties are incorrect. The question to be answered is: How robust are the estimates to the utilization of incorrect prior knowledge, and how do correlations between independent variables affect this robustness? In highly correlated cases, the addition of prior knowledge modifies properties of estimates in a way that might lead the experimenter to use any prior information at his disposal; if robustness is evident, imperfection of prior knowledge should cause less concern.

2.2. Maximum Likelihood Estimators and Their Properties Using No Priors or Correct Priors

As is known, when no prior information is used, and the model is

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon},\ \underline{\varepsilon} \sim MN(\underline{0};\sigma^2 I_n)\ , \tag{2.14}$$

it follows that

$$\hat{\underline{\beta}} = S^{-1}X'\underline{y};\ \hat{\sigma}^2 = (\underline{y} - X\hat{\underline{\beta}})'(\underline{y} - X\hat{\underline{\beta}})/(n-p). \tag{2.15}$$

In the above, y has n elements, X is n by p, S is p by p, and $\hat{\underline{\beta}}$ has p elements. Properties of these estimators are:

$$\begin{aligned} E\hat{\underline{\beta}} &= \underline{\beta};\quad E\hat{\sigma}^2 = \sigma^2; \\ V\hat{\underline{\beta}} &= MSE\hat{\underline{\beta}} = S^{-1}\sigma^2;\ V\hat{\sigma}^2 = MSE\hat{\sigma}^2 = 2\sigma^4/(n-p); \\ V\hat{\underline{y}}_0 &= X_0 S^{-1} X_0' \sigma^2\ , \end{aligned} \tag{2.16}$$

where X_0 specifies the set of points at which $\hat{\underline{y}}$ is calculated.

When prior information is added, the model (2.1) can be rewritten as

$$\underline{y}^* = X^*\underline{\beta} + \underline{\varepsilon}^*,\ \underline{\varepsilon}^* \sim MN(\underline{0};\sigma^2 I_{n+p}) \ ; \tag{2.17}$$

where $\underline{y}^* = \begin{bmatrix} \underline{y} \\ D^{-1/2}\underline{\beta}_0 \end{bmatrix}$, $X^* = \begin{bmatrix} X \\ D^{-1/2} \end{bmatrix}$, and $\underline{\varepsilon}^* = \begin{bmatrix} \underline{\varepsilon} \\ D^{-1/2}\underline{\delta} \end{bmatrix}$.

Using this model, we have[4]:

$$\tilde{\underline{\beta}} = \tilde{S}^{-1}X^{*\prime}\underline{y}^*;\quad \tilde{\sigma}^2 = (\underline{y}^*-X^*\tilde{\underline{\beta}})'(\underline{y}^*-X^*\tilde{\underline{\beta}})/n \ . \tag{2.18}$$

$$E\tilde{\underline{\beta}} = \underline{\beta};\ E\tilde{\sigma}^2 = \sigma^2;\ V\tilde{\underline{\beta}} = MSE\tilde{\underline{\beta}} = \tilde{S}^{-1}\sigma^2;$$

$$V\tilde{\sigma}^2 = MSE\tilde{\sigma}^2 = 2\sigma^4/n;\ V\tilde{\underline{y}}_0 = X_0\tilde{S}^{-1}X_0'\sigma^2 \ . \tag{2.19}$$

When knowledge is available on σ^2, and model (2.1) is used with model (2.2), estimates and their properties are as follows[5]. When (c/σ^2) is distributed as a χ^2 variate with k degrees of freedom,

$$\tilde{\underline{\beta}} = \tilde{S}^{-1}X^{*\prime}\underline{y}^*;\ \tilde{\sigma}_c^2 = [(\underline{y}^*-X^*\tilde{\underline{\beta}})'(\underline{y}^*-X^*\tilde{\underline{\beta}})+c]/(n+k) \ . \tag{2.20}$$

$$E\tilde{\underline{\beta}} = \underline{\beta};\ E\tilde{\sigma}_c^2 = \sigma^2;\ V\tilde{\underline{\beta}} = \tilde{S}^{-1}\sigma^2;$$

$$V\tilde{\sigma}_c^2 = 2\sigma^4/(n+k);\ V\tilde{\underline{y}}_0 = X_0\tilde{S}^{-1}X_0'\sigma^2 \ . \tag{2.21}$$

2.3. Properties When the Prior Information Is Incorrect

2.3.1. Normally Distributed Prior Densities

The experimenter, in obtaining estimators (2.18), believes his prior knowledge, $\underline{\beta}_0$, to be distributed with mean $\underline{\beta}$, and variance $D\sigma^2$.

---

[4]Theorems required for the results presented here are given in Appendix 7.3.

[5]As noted in Appendix 7.1, when prior knowledge on σ^2 is used, $E(c) = k\sigma^2$ and $V(c) = 2k\sigma^4$.

If the prior has mean $\underline{\beta} + \underline{m}$, and variance $D_0\sigma^2$, the correct distribution of the y* vector is

$$\underline{y}^* \sim MN(\underline{u};V\sigma^2), \tag{2.22}$$

where

$$\underline{u} = \begin{bmatrix} X\underline{\beta} \\ D^{-1/2}(\underline{\beta}+\underline{m}) \end{bmatrix} \quad ; \qquad V = \begin{bmatrix} I_n & 0_{np} \\ 0_{pn} & D_0D^{-1} \end{bmatrix} .$$

Then, the correct properties are, using the theorems of Appendix 7.3,

$$E_w\tilde{\underline{\beta}} = \underline{\beta} + \tilde{S}^{-1}D^{-1}\underline{m};$$

$$E_w\tilde{\sigma}^2 = [\underline{u}'A\underline{u} + \sigma^2 Tr(VA)]/n ,$$

where $A = I_{n+p} - X^*\tilde{S}^{-1}X^{*\prime}$, and Tr(VA) is the trace of the matrix VA;

$$\begin{aligned}
V_w\tilde{\underline{\beta}} &= \tilde{S}^{-1}X^{*\prime}VX^*\tilde{S}^{-1}\sigma^2; \\
V_w\tilde{\sigma}^2 &= [4\underline{u}'AVA\underline{u}\sigma^2 + 2\sigma^4 Tr(VA)^2]/n^2; \\
MSE_w\tilde{\underline{\beta}} &= V_w\tilde{\underline{\beta}} + \tilde{S}^{-1}D^{-1}\underline{m}\underline{m}'D^{-1}\tilde{S}^{-1}; \qquad (2.23) \\
MSE_w\tilde{\sigma}^2 &= V_w\tilde{\sigma}^2 + (E_w\tilde{\sigma}^2 - \sigma^2)^2; \\
V_w\tilde{\underline{y}}_0 &= X_0\tilde{S}^{-1}(S + D^{-1}D_0D^{-1})\tilde{S}^{-1}X_0'\sigma^2,
\end{aligned}$$

where X_0 is the set of points at which $\tilde{\underline{y}}$ is calculated.

2.3.2. Normally Distributed Incorrect Prior Densities on the Parameters $\underline{\beta}$, and Incorrect Priors on σ^2

The experimenter may have used model (2.1) with model (2.2) to obtain estimators (2.20) with expected properties (2.21). If c/σ^2 is actually distributed as a chi-square variate with k_0 degrees of freedom, rather than k, while $\underline{\beta}_0$ is incorrect as specified above in Section 2.3.1, the properties of estimators (2.20) are

$$E_w\tilde{\sigma}_c^2 = [\underline{u}'A\underline{u} + \sigma^2 Tr(VA) + k_0\sigma^2]/(n+k);$$
$$V_w\tilde{\sigma}_c^2 = [4\underline{u}'AVA\underline{u}\sigma^2 + 2\sigma^4 Tr(VA)^2 + 2k_0\sigma^4]/(n+k)^2 . \quad (2.24)$$

A and B are defined in equation (2.23). Other properties are unchanged from those given in equations (2.23).

If c/σ^2 is not chi-square in form, but exponential[6], then

$$E_w\tilde{\sigma}_c^2 = [\underline{u}'A\underline{u} + \sigma^2 Tr(VA) + k_0\sigma^2]/(n+k);$$
$$V_w\tilde{\sigma}_c^2 = [4\underline{u}'AVA\underline{u}\sigma^2 + 2\sigma^4 Tr(VA)^2 + k_0^2\sigma^4]/(n+k)^2 . \quad (2.25)$$

2.3.3. <u>Non-normally Distributed Prior Densities</u>

In this dissertation, the errors in prior specification made by the experimenter are considered to be mean and variance specification errors. All non-normal distributions considered, i.e., the beta, uniform, and exponential, are given the same first and second moments as were given to the incorrect normal distributions, discussed in section 2.3.1. This means that properties (2.23) of $\tilde{\underline{\beta}}$, (2.18), for the normal distribution apply when the distributions are non-normal.

When estimation of variance is considered, third and fourth moments are important. For the distributions applied here, third

---

[6]This prior density is developed in Appendix 7.1.5; $E(c) = k_0\sigma^2$ and $V(c) = k_0^2\sigma^4$.

moments of the beta and uniform densities are zero, but fourth moments differ from those of the normal distribution. For the exponential distribution, both third and fourth moments differ. In this dissertation, the estimation of variance is discussed only for normally distributed prior situations.

Although the means and variances of estimators of $\underline{\beta}$ are not modified by non-normality, and hence not considered in the calculations, non-normality should affect power. This problem is discussed in Chapter 3.

An area for future investigation which could be considered is the application of non-normally distributed priors to the regression model, obtaining estimates and properties, and viewing these properties when the assumed priors are incorrect. If beta priors were added, the normal equations would be (using equation 7.1):

$$\frac{\partial \log L}{\partial \beta_j} = (\underline{y}'\underline{x}_j - \underline{\beta}'\underline{s}_j)/\sigma^2 - \frac{(\theta_{1j}-1)}{\beta_{0j}-\beta_j-c_{1j}} - \frac{(\theta_{2j}-1)}{c_{2j}-\beta_{0j}+\beta_j} \;;$$

$j = 1,\ldots,p$; $\underline{x}_j$ and $\underline{s}_j$ are the jth columns of X and $S = X'X$, respectively. The θ's are the parameters of the beta density, and the c's are its endpoints. (2.26)

$$\frac{\partial \log L}{\partial \sigma^2} = -n/2\sigma^2 + (\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta})/2\sigma^4.$$

Defining $\underline{\omega}' = (\underline{\beta}', \sigma^2)$, $\partial \log L/\partial \omega_i = 0 = Z(\underline{\omega})$, $i = 1,\ldots,p+1$, must be solved by iterative methods. Expanding $Z(\underline{\omega})$ in a Taylor Series about an initial guess $\underline{\omega}_0$, the iterative scheme

$$\underline{\omega}_{i+1} = \underline{\omega}_i - [Z'(\underline{\omega}_i)]^{-1} Z(\underline{\omega}_i) \tag{2.27}$$

results from neglecting second and higher order derivatives, and could be used. Its use is discussed in detail in Appendix 7.4.

When the prior knowledge added is exponential, similar, but simpler, equations result. Appendix 7.4 also gives this development.

2.3.4. Further Discussion

The computational results of using the developments given in this chapter are displayed in Chapter 4. Chapter 4 discusses two programs and their results: (1) A calculation program based on the current chapter; and (2) A simulation program based on the priors discussed here and the development given in the next chapter.

In Chapter 4, tables of $MSE_w\tilde{\beta}_1$ and $V\hat{\beta}_1$ are given for various values of X, the regression matrix; D, the assumed prior variance; D_0, the correct prior variance[7]; and $\underline{m}$, the bias in the mean of the prior knowledge. These tables can be used to determine the conditions under which the use of prior information, even though biased, would still produce estimates of $\underline{\beta}$ superior to $\underline{\hat{\beta}}$. Tables are also displayed pertaining to estimates of variance, with and without the use of prior information on σ^2.

---

[7]In the programs described in Chapter 4, σ^2 is given unit value. Thus, although the assumed and true variances of $\underline{\delta}$ and $D\sigma^2$ and $D_0\sigma^2$, hereafter in this thesis these variances are given the values D and D_0 unless clarity dictates otherwise.

3. THE POWER OF TESTS WHEN USING PRIOR INFORMATION

This chapter will study estimation using prior knowledge by viewing the power of tests of hypotheses. The model used will be, as before,

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon},\ \underline{\varepsilon} \sim MN(\underline{0},\ \sigma^2 I_n);$$
$$\underline{\beta}_0 = \underline{\beta} + \underline{\delta},\ \underline{\delta} \sim MN(\underline{0};\ \sigma^2 D)\ , \qquad (3.1)$$

and the hypothesis under investigation will be

$$H_0:\ \underline{\beta}_1 = \underline{0},\ \text{versus}\ H_a:\ \underline{\beta}_1 = \underline{\beta}_1^* \ . \qquad (3.2)$$

$\underline{\beta}_1$ has $r \leq p$ elements, and the nullity of H_0 is without loss of generality. The various models considered will allow judgment to be made on the general effect conditioning has on power.

Two cases will be investigated: The powers

a. $\P(\tilde{\lambda})$, where $\tilde{\lambda}$ is the non-centrality parameter associated with with $\underline{\beta}_1^*$, when prior knowledge (3.1) is used to find $\underline{\tilde{\beta}}$. $\P(\tilde{\lambda})$ is the power that the experimenter thinks he has achieved, at some level α.

b. $\P(\tilde{\lambda}_w)$, the power of the actual test, where $\underline{\tilde{\beta}}$ was used, but prior knowledge was incorrect.

The question answered will be, than: If an experimenter thinks he is testing H_0 at level α, and obtaining power $\P(\tilde{\lambda})$, and if his prior knowledge is actually wrong, what are the true α and $\P(\tilde{\lambda})$, termed α_w and $\P(\tilde{\lambda}_w)$?

Another situation is also considered. A. E. Hoerl (1959) has developed an algorithm, based on his work in ridge analysis of response surfaces, which may be useful in the regression problem when highly correlated independent variables are involved. Mr. Hoerl does not obtain best linear unbiased estimators in his work; rather, he minimizes another quantity termed ERR, within certain restrictions, where ERR is simply $\sigma^2 Tr(X'X)^{-1}$ for the usual least squares estimator of $\underline{\beta}$. In an ill conditioned situation, the variances of BLU estimators are quite large; that is, $Tr(X'X)^{-1}$ is large. If a well-founded technique can be used to decrease this variance drastically while producing a small bias, then this technique may be of practical importance.

In his monograph, now being prepared for publication, Hoerl chooses a conditioning parameter, 1/d, which is added to each of the diagonal elements of X'X. This parameter comes from minimizing ERR(1/d) under restraints; the choice of 1/d is dependent on the condition of X'X. Thus, Hoerl's estimator is the same as the estimator $\hat{\underline{\beta}}$(2.18) when the X variables have been standardized to have unit variance, when the prior information is null, and when $D = dI_p$. This means that

$$E_w \underline{\beta}_h = \underline{\beta} - \tilde{S}^{-1} D^{-1} \underline{\beta} ,$$

since prior knowledge, in the Hoerl framework[8], is biased by $-\underline{\beta}$.

The consequences of using incorrect prior knowledge are important in the application of a conditioning parameter, and power modifications which come from using Hoerl's estimation scheme are discussed below in the framework of prior knowledge.

---

[8] $\underline{\beta}_h$, Hoerl's estimator of $\underline{\beta}$, is $\underline{\beta}_h = \tilde{S}^{-1} X'\underline{y}$.

Partitioning as dictated by H_0, the model (2.17) is

$$\underline{y}^* = X_1^*\underline{\beta}_1 + X_2^*\underline{\beta}_2 + \underline{\varepsilon}^*$$

$$= \begin{bmatrix} X_1 \\ D_1^{-1/2} \\ 0 \end{bmatrix} \underline{\beta}_1 + \begin{bmatrix} X_2 \\ 0 \\ D_2^{-1/2} \end{bmatrix} \underline{\beta}_2 + \underline{\varepsilon}^*; \quad D^{-1/2} = \begin{bmatrix} D_1^{-1/2} & 0 \\ & \\ 0 & D_2^{-1/2} \end{bmatrix} . \quad (3.3)$$

The dimensions of the matrices given in (3.3) are: (n by r) for X_1; (r by r) for D_1; [n by (p-r)] for X_2; and [(p-r) by (p-r)] for D_2.

If no prior knowledge has been added to the model, the statistic used to test H_0 is

$$\frac{(n-p)}{r} \frac{[SS_0 - SSE]}{SSE} + \frac{(n-p)[\hat{\underline{\beta}}'X'\underline{y} - \tilde{\hat{\underline{\beta}}}_2' X_2' \underline{y}]}{r \quad [\underline{y}'\underline{y} - \hat{\underline{\beta}}'X'\underline{y}]} \overset{H_0}{=} F(r, n-p), \quad (3.4)$$

where SSE_0 is the error sum of squares under H_0, SSE is the error sum of squares under the full model, and $\tilde{\hat{\underline{\beta}}}_2$ is the estimator for $\underline{\beta}_2$ when $\underline{\beta}_1 = \underline{0}$;

$$\tilde{\hat{\underline{\beta}}}_2 = (X_2'X_2)^{-1} X_2'\underline{y} . \quad (3.5)$$

The non-centrality parameter, $\hat{\lambda}$, associated with this statistic, is

$$\hat{\lambda} = \underline{\beta}_1'[X_1'X_1 - X_1'X_2(X_2'X_2)^{-1}X_2'X_1]\underline{\beta}_1/(2\sigma^2) . \quad (3.6)$$

When prior knowledge is added, as in (3.1), the statistic and non-centrality parameter, $\tilde{\lambda}$, are rewritten using (3.4), (3.5), and (3.6), with X* replacing X, and $\tilde{\underline{\beta}}$ replacing $\hat{\underline{\beta}}$.

The increase in power when correct prior knowledge is added to the model can be seen by noting that $2\sigma^2(\tilde{\lambda}-\hat{\lambda})$ is always positive, and that power is monotone increasing in the non-centrality parameter[9]. Actually,

$$2\sigma^2(\tilde{\lambda}-\hat{\lambda}) = \underline{\beta}_1' A \underline{\beta}_1 \ , \tag{3.7}$$

and A is positive definite when D_2^{-1} is non-null:

$$A = D_1^{-1} + X_1'X_2[(X_2'X_2)^{-1} - (X_2'X_2 + D_2^{-1})^{-1}]X_2'X_1 . \tag{3.8}$$

When the experimenter uses incorrect normally distributed prior knowledge, the actual power calculations are more complex. Statistic (3.4) can be used to test H_0, in which X is replaced by X*, $\underline{y}$ by $\underline{y}^*$, $\hat{\underline{\beta}}$ by $\tilde{\underline{\beta}}$, and n-p by n. This can be rewritten as

$$F^* = \frac{n(\underline{y}^{*\prime} A_1 \underline{y}^*)}{r(\underline{y}^{*\prime} A_2 \underline{y}^*)} \ , \tag{3.9}$$

where $A_1 = X^*(X^{*\prime}X^*)^{-1}X^{*\prime} - X_2^*(X_2^{*\prime}X_2^*)^{-1}X_2^{*\prime}$, and

$$A_2 = I_{n+p} - X^*(X^{*\prime}X^*)^{-1}X^{*\prime} \ .$$

In (3.9), $\underline{y}^{*\prime} = (\underline{y}' ; \underline{\beta}_0' D^{-1/2})$ was used by the experimenter to obtain estimates. He believes that the distribution of the two parts of $\underline{y}^*$ are

$$\underline{y} \sim MN(X\underline{\beta};\ \sigma^2 I_n)$$

$$D^{-1/2}\underline{\beta}_0 \sim MN(D^{-1/2}\underline{\beta};\ \sigma^2 I_p). \tag{3.10}$$

---

[9] Course notes, Dr. R. C. Bose, Analysis of Variance with Application to Experimental Designs, University of North Carolina, Chapel Hill, N. C., page 65.

However, the true situation is that

$$\underline{y}^* \sim MN(\underline{u};\ V\sigma^2), \tag{3.11}$$

where $\underline{u}' = [\underline{\beta}'X';\ (\underline{m}+\underline{\beta})'D^{-1/2}]$,

and

$$V = \begin{bmatrix} I_n & 0_{np} \\ 0_{pn} & D^{-1}D_0 \end{bmatrix} .$$

Under these conditions, the statistic F* in (3.9) can be rewritten as follows:

$$F^* = \frac{n[(X^*\underline{\beta} + G\underline{m} + \underline{\nu})'A_1(X^*\underline{\beta} + G\underline{m} + \underline{\nu})]}{r[(X^*\underline{\beta} + G\underline{m} + \underline{\nu})'A_2(X^*\underline{\beta} + G\underline{m} + \underline{\nu})]} , \tag{3.12}$$

where $\underline{\nu} \sim MN(\underline{0}_{n+p};\ V\sigma^2)$;

$$G = \begin{bmatrix} 0_{np} \\ D^{-1/2} \end{bmatrix} .$$

This is equivalent to

$$F^* = \frac{n[\phi(\underline{\beta}_1) + (\underline{\nu} + G\underline{m})'A_1(\underline{\nu} + G\underline{m})]}{r[(\underline{\nu} + G\underline{m})'A_2(\underline{\nu} + G\underline{m})]} , \tag{3.13}$$

where $\phi(\underline{\beta}_1) = \underline{\beta}_1'X_1^{*\prime}A_3X_1^*\underline{\beta}_1 + 2\underline{\beta}_1'X_1^{*\prime}A_3G\underline{m} + 2\underline{\beta}_1X_1^{*\prime}A_3\underline{\nu}$, with

$$A_3 = I - X_2^*(X_2^{*\prime}X_2^*)^{-1}X_2^{*\prime} .$$

Elements of F* involving $\underline{\beta}_2$ directly are eliminated because

$A_2X^* = 0$, and

$$X^{*\prime}A_1X^* = \begin{bmatrix} X_1^{*\prime}A_3X_1^* & 0 \\ 0 & 0 \end{bmatrix} . \tag{3.14}$$

Scrutiny of (3.13) brings out the following:

1. Since $A_1A_2=0$, then, when $\underline{m}$ is null, F* is distributed as the non-central F, $F^{**}(r,n,\lambda)$, where $\lambda = 0$ under H_0.

2. When $\underline{m}$ is non-null, an approximate F statistic can be derived. (Imhof, 1961 ; Harvey, 1965.)

In this chapter, we are also considering the power of the test H_0: $\underline{\beta}_1 = \underline{0}$ when the pertinent quadratic forms involve non-normally distributed y* vectors. Simulations are required for these power investigations; therefore this thesis will also simulate the case of normally distributed incorrect prior information, rather than use an approximate F statistic for power evaluation.

A computer program, described in Section 4.2 of the following chapter, was designed to consider the following problem and related questions.

Given the model

$$\begin{aligned} \underline{y} &= \beta_1\underline{x}_1 + \beta_2\underline{x}_2 + \underline{\varepsilon},\ \underline{\varepsilon} \sim MN(\underline{0};\ \sigma^2 I_n), \\ \beta_{01} &= \beta_1 + \delta_1,\ \delta_1 \sim N(0;\ d_1\sigma^2), \\ \beta_{02} &= \beta_2 + \delta_2,\ \delta_2 \sim N(0;\ d_2\sigma^2), \end{aligned} \qquad (3.15)$$

generate the above random variables, and determine $\P(\hat{\lambda})$, the "no prior" power, and $\P(\tilde{\lambda})$, the power when prior knowledge is added. Then determine $\P(\tilde{\lambda}_w)$, the power when incorrect prior knowledge is used, by generating $\underline{\delta}$'s from the normal, beta, uniform, and exponential distributions as described in Chapter 2, equations (2.3) through (2.6).

The hypothesis considered was H_0: $\beta_1 = 0$, and the statistics used were:

$$(1) \quad \frac{\hat{\underline{\beta}}'X'\underline{y} - \hat{\beta}_2 x_2'\underline{y}}{[\underline{y}'\underline{y} - \hat{\underline{\beta}}'X'\underline{y}]/(n-2)} \quad \text{for } \P(\hat{\lambda}), \text{ where} \qquad (3.16)$$

$$\hat{\beta}_2 = (\underline{x}_2'\underline{x}_2)^{-1}\underline{x}_2'\underline{y} \ ;$$

$$(2) \quad \frac{\tilde{\underline{\beta}}'X^{*\prime}\underline{y}^* - \tilde{\beta}_2\underline{x}_2^{*\prime}y^*}{[\underline{y}^{*\prime}\underline{y}^* - \tilde{\underline{\beta}}'X^{*\prime}\underline{y}^*]/n} \quad \text{for } \P(\tilde{\lambda}), \text{ where} \qquad (3.17)$$

$$\tilde{\beta}_2 = (\underline{x}_2^{*\prime}\underline{x}_2^*)^{-1}\underline{x}_2^{*\prime}\underline{y}^* \ .$$

For $\P(\tilde{\lambda}_w)$, statistic (2) is used, but $\underline{y}^{*\prime} = (\underline{y}', \underline{\beta}_0'D^{-1/2})$ is generated using correct prior values.

$$\underline{y}^{*\prime} = [\underline{\beta}'X' + \underline{\varepsilon}'; (\underline{m}+\underline{\beta})'D^{-1/2} + \underline{\delta}'D^{-1/2}] \text{ where } V(\underline{\varepsilon}) = \sigma^2 I_n, \text{ and } V(\underline{\delta}) = \sigma^2 D_0.$$

Also, statistic (2) is calculated using $\underline{\beta}_0 = \underline{0}$, so that the estimator used by Hoerl can be evaluated.

In the program, the models below were used, so that the size, and the power at two different levels of β_1, could be determined. The models were:

$$\underline{y} = \underline{\varepsilon}; \qquad \underline{y} = \underline{x}_1 + \underline{\varepsilon}; \qquad \underline{y} = 5\underline{x}_1 + \underline{\varepsilon};$$

$$\underline{y} = \underline{x}_2 + \underline{\varepsilon}; \quad \underline{y} = \underline{x}_1 + \underline{x}_2 + \underline{\varepsilon}; \quad \underline{y} = 5x_1 + x_2 + \underline{\varepsilon}. \qquad (3.19)$$

Ten observations were used, and when the statistic exceeded the appropriate (8 or 10 degrees of freedom) t-value, a count was made. σ^2 was given a value of one, and assumed prior variances of $\underline{\delta}$,

$D = 1000I$, $100I$, and $10I$, were used. That is, in (3.15), $d_1 = d_2 = d$ equal to 1000, 100, and 10 were the variances used. Correct prior variances, d_0, were taken to be $d/5$, d, and $5d$, where $D_0 = d_0 I$. Densities of priors were taken to be normal, beta, exponential, and uniform.

That is, knowledge was considered to be incorrect in the above combinations of its variance and distributional form, for each model simulated, for every correlation considered. Two hundred fifty simulations were run.

In addition, one hundred simulations were computed as described above, using means, $\underline{m}$, of $\sqrt{d}$ and $2\sqrt{d}$. This allowed a determination of the effect of bias on power. Special cases using means $\underline{m}'$ of (1,1) and (2,2) were also simulated.

4. COMPUTER PROGRAMS AND RESULTS

This chapter is divided into two sections. The first discusses the calculation of properties, not including power, of estimators. Explicit results, given in Chapter 2, are available for calculations. The second section discusses simulations which allow power determination. In those simulations, the mean square error and the average of the estimates are also calculated. This allows a check between the two programs.

The two computer programs consider that prior knowledge is added with variance $D = dI$. Although this would not generally be the situation, two factors were considered which led to this choice.

1. Hoerl's estimation technique involved adding a constant to the diagonal of $X'X$; the analogy with the addition of prior knowledge requires that $D = dI$. This choice allows consideration of how Hoerl's estimator modifies power of tests of hypotheses (Section 4.2.6), and what effect his technique has on the estimation of parameters of varying importance (Section 4.1.1).

2. An overview of the relationship between adding incorrect prior knowledge and ill-condition will come from such a choice; specific areas for further study will be defined.

In the programs, σ^2 is given unit value; therefore $D\sigma^2$, the variance-covariance matrix of the vector $\underline{\delta}$, is written as D. $D = dI$ in these calculations; i.e., $d_1 = d_2 = d$ in (3.15), for the simulation program, as well as for the program described in the following section.

4.1. Properties of Estimators and Prior Knowledge

Figure 4.2 outlines a computer program designed to show how the mean square errors of estimates of $\underline{\beta}$ and σ^2 change with correlation. The program accepts a general model, and calculates the mean square errors for various situations. Models with two parameters were examined by the program. Only normally distributed correct priors are considered. Results are in terms of estimators of β_1 and σ^2.

Certain of the calculations involve the matrix X. The four X matrices required were computed as follows, so as to give the desired correlation matrices. First, orthogonal vectors were calculated, as described in Anderson and Bancroft (1952), Chapter 16. These vectors were normalized, so that the resultant X'X matrix was an identity matrix. Perturbations on the outlying elements of these vectors, and normalization of the resultant vectors, gave vectors such that X'X was the required correlation matrix. Figure 4.1 shows the range of X as produced by the above calculations, for all correlations considered.

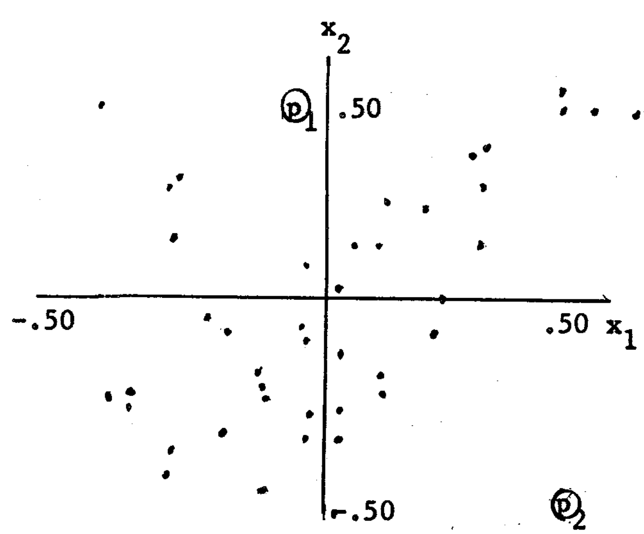

Figure 4.1. The range of X

- All prior variances, d
- Correct prior variances, D_0
- Correct prior means, $\underline{m}$
- c/σ^2 degrees of freedom, k
- Correct degrees of freedom, k_0
- Number of observations
- Number of parameters, $\underline{\beta}$
- Variance used, σ^2
- X_0, points at which $V\hat{\underline{y}}_0$ is calculated
- X matrices

<u>List A</u>

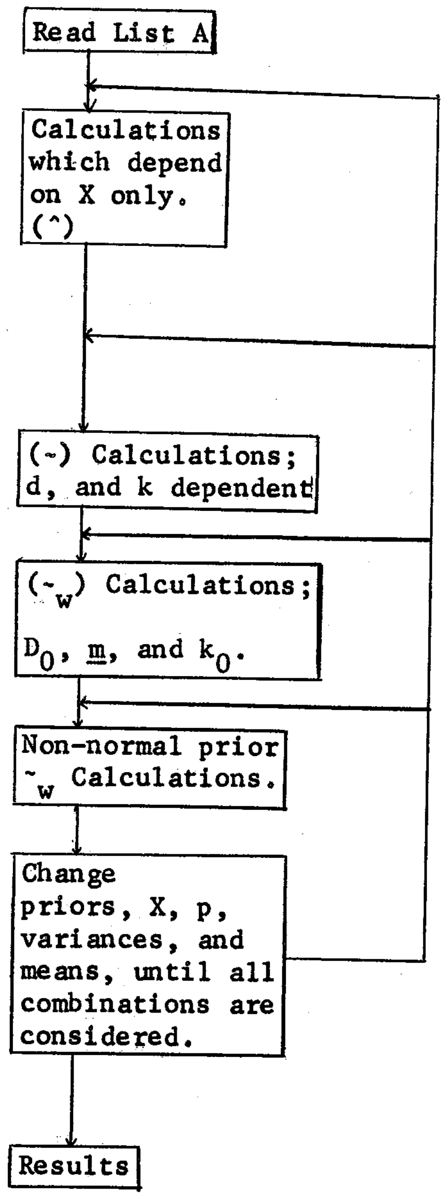

Figure 4.2. The program which calculates properties of estimates

The concern of this investigation is to determine the interplay between:

1. The condition of X'X, implied by the magnitude of ρ_{ij},
2. The addition of prior knowledge with variance D, and
3. The incorrectness of that prior knowledge, in terms of
   a. Variance used, D, the actual variance being D_0,
   b. Mean used, $\underline{0}$, the actual mean being $\underline{m}$.

For models (2.1), and (2.1) with (2.2), all combinations of the following were considered in calculations. Considerable investigation and calculation was required to determine parameter ranges of interest. The ranges given were used in program one, and the interesting subset of these ranges is tabulated in tables in this section. In program two, the simulation program described in Section 4.2, ranges used for $\underline{m}$ and D = dI were slightly modified from those used here.

1. $\rho = (0, .5, .85, .96)$ (correct to two decimal places)
2. $d = (1000, 100, 10, 1, 1/2)$; $dI = D$
3. $D_0 = (D/5, D, 5D)$; $d_0 I = D_0$
4. $\underline{m} = (\underline{0}, \underline{.5\sqrt{d}}, \underline{\sqrt{d}}, \underline{2\sqrt{d}})$.
5. Normal prior knowledge, with means $\underline{m}$ and variance D_0, was used.

When prior knowledge on σ^2 was added, the following combinations were added to the above:

6. $k = (8, 16, 32)$ (degrees of freedom for χ^2)
7. $k_0 = (k + 4)$ (correct degrees of freedom)
8. χ^2 prior knowledge was used, with χ^2 and exponential distributions as the correct priors.

Program results are given in the following sections.

4.1.1. Mean Square Error of $\underline{\beta}_1$

Tables 4.1 through 4.3 show how the mean square error of the estimator of β_1 is changed by the use of prior knowledge. The degree to which $MSE\hat{\beta}_1$, which is also $V\hat{\beta}_1$, is affected depends on the magnitude of

1. The assumed prior variance, D, where D = dI;
2. The correlations between the independent variables, ρ_{ij}.

Table 4.1 gives results in terms of actual values of the variance. Tables 4.2 and 4.3 display ratios of mean square errors, so as to facilitate inferences about the following:

1. The effect of bias in the mean of the prior, and its relationship to prior variance used and correlation;
2. The effect of an error in choosing the variance of the prior.

The effect of using correct prior knowledge can be seen by viewing Table 4.1 below.

Table 4.1. $V\hat{\beta}_1$ and $V\tilde{\beta}_1$; using correct prior knowledge on $\underline{\beta}$

	Correlation				
	0	.50	.85	.96	d
$V\hat{\beta}_1$	1.00	1.33	3.62	13.31	
$V\tilde{\beta}_1$	1.00	1.33	3.60	12.98	1000.0
	0.99	1.31	3.41	10.61	100.0
	0.91	1.15	2.26	3.86	10.0
	0.50	0.53	0.61	0.65	1.0
	0.33	0.34	0.36	0.37	0.5

Table 4.2. $MSE_w\tilde{\beta}_1/V\hat{\beta}_1$; $MSE_w\tilde{\beta}_1/V\tilde{\beta}_1$; biases in the prior knowledge on $\underline{\beta}$

	Correlation				
$\underline{m}\sqrt{d}$	0	.50	.85	.96	d
.00	1.00;1.00	1.00;1.00	0.99;1.00	0.98;1.00	1000.0
.50	1.00;1.00	1.00;1.00	0.99;1.00	0.98;1.00	1000.0
1.00	1.00;1.00	1.00;1.00	0.99;1.00	0.98;1.00	1000.0
2.00	1.00;1.00	1.00;1.00	0.99;1.00	0.98;1.00	1000.0
.00	0.99;1.00	0.98;1.00	0.94;1.00	0.80;1.00	100.0
.50	0.99;1.00	0.98;1.00	0.94;1.00	0.80;1.00	100.0
1.00	1.00;1.01	0.99;1.00	0.94;1.00	0.80;1.00	100.0
2.00	1.03;1.04	1.00;1.01	0.95;1.00	0.80;1.00	100.0
.00	0.91;1.00	0.86;1.00	0.63;1.00	0.29;1.00	10.0
.50	0.93;1.02	0.87;1.01	0.63;1.00	0.29;1.00	10.0
1.00	0.99;1.09	0.89;1.03	0.63;1.01	0.29;1.01	10.0
2.00	1.24;1.36	0.98;1.14	0.65;1.05	0.30;1.02	10.0
.00	0.50;1.00	0.40;1.00	0.17;1.00	0.05;1.00	1.0
.50	0.56;1.13	0.43;1.08	0.18;1.05	0.05;1.04	1.0
1.00	0.75;1.50	0.52;1.30	0.20;1.20	0.06;1.18	1.0
2.00	1.50;3.00	0.88;2.20	0.31;1.81	0.08;1.70	1.0
.00	0.33;1.00	0.26;1.00	0.10;1.00	0.03;1.00	0.5
.50	0.39;1.17	0.29;1.12	0.11;1.09	0.03;1.09	0.5
1.00	0.56;1.67	0.38;1.48	0.14;1.37	0.04;1.34	0.5
2.00	1.22;3.67	0.75;2.91	0.25;2.49	0.07;2.37	0.5

Table 4.3. $MSE_w\tilde{\beta}_1/V\hat{\beta}_1$; $MSE_w\tilde{\beta}_1/V\tilde{\beta}_1$; errors in the mean and variance of the prior on $\underline{\beta}$

$md^{-1/2}$	Correlation 0	.50	.85	.96	d
			$d = 5d_0$		
0.0 to 2.0	1.00;1.00	1.00;1.00	0.99;1.00	0.96;0.98	1000.0
0.0	0.98;0.99	0.97;0.99	0.90;0.95	0.67;0.84	100.0
0.5	0.99;1.00	0.97;0.99	0.90;0.95	0.67;0.84	100.0
1.0	0.99;1.00	0.97;0.99	0.90;0.95	0.67;0.84	100.0
2.0	1.02;1.03	0.98;1.00	0.90;0.96	0.67;0.84	100.0
0.0	0.84;0.93	0.76;0.89	0.44;0.71	0.13;0.46	10.0
0.5	0.86;0.95	0.77;0.90	0.45;0.71	0.13;0.46	10.0
1.0	0.93;1.02	0.79;0.92	0.45;0.72	0.14;0.46	10.0
2.0	1.17;1.29	0.88;1.03	0.47;0.76	0.14;0.48	10.0
0.0	0.30;0.60	0.22;0.55	0.07;0.42	0.02;0.36	1.0
0.5	0.36;0.73	0.25;0.62	0.08;0.47	0.02;0.40	1.0
1.0	0.55;1.10	0.34;0.85	0.11;0.63	0.03;0.54	1.0
2.0	1.30;2.60	0.70;1.75	0.21;1.23	0.05;1.06	1.0
0.0	0.16;0.47	0.11;0.44	0.04;0.37	0.01;0.34	0.5
0.5	0.21;0.63	0.14;0.56	0.05;0.47	0.01;0.43	0.5
1.0	0.38;1.13	0.24;0.91	0.08;0.75	0.02;0.69	0.5
2.0	1.04;3.13	0.60;2.34	0.19;1.86	0.05;1.72	0.5
			$d = d_0/5$		
0.0	1.00;1.00	1.01;1.01	1.02;1.03	1.07;1.10	1000.0
0.5	1.00;1.00	1.01;1.01	1.02;1.03	1.07;1.10	1000.0
1.0	1.00;1.01	1.01;1.01	1.02;1.03	1.07;1.10	1000.0
2.0	1.01;1.01	1.01;1.01	1.02;1.03	1.07;1.10	1000.0
0.0	1.03;1.04	1.05;1.07	1.16;1.23	1.44;1.81	100.0
0.5	1.03;1.04	1.05;1.07	1.16;1.23	1.44;1.81	100.0
1.0	1.04;1.05	1.05;1.07	1.16;1.23	1.44;1.81	100.0
2.0	1.07;1.08	1.06;1.08	1.16;1.24	1.44;1.81	100.0
0.0	1.24;1.36	1.34;1.55	1.53;1.45	1.08;3.72	10.0
0.5	1.26;1.39	1.34;1.56	1.53;2.45	1.08;3.72	10.0
1.0	1.32;1.46	1.36;1.59	1.54;2.46	1.08;3.73	10.0
2.0	1.57;1.73	1.45;1.69	1.56;2.49	1.09;3.75	10.0

Table continued

Table 4.3. (continued)

$md^{-1/2}$	Correlation 0	.50	.85	.96	d
			$d = d_0/5$		
0.0	1.50;3.00	1.31;3.27	0.66;3.88	0.21;4.20	1.0
0.5	1.56;3.13	1.34;3.34	0.66;3.93	0.21;4.25	1.0
1.0	1.75;3.50	1.43;3.57	0.69;4.09	0.21;4.39	1.0
2.0	2.50;5.00	1.79;4.47	0.79;4.69	0.24;4.90	1.0
0.0	1.22;3.67	0.98;3.82	0.41;4.13	0.12;4.28	0.5
0.5	1.27;3.83	1.01;3.94	0.42;4.23	0.12;4.36	0.5
1.0	1.44;4.33	1.11;4.30	0.45;4.51	0.13;4.62	0.5
2.0	2.11;6.33	1.47;5.72	0.56;5.62	0.16;5.65	0.5

The use of correct prior knowledge provides the experimenter with advantages which are easily seen. The effect of such model augmentation is dependent on correlations between the independent variables, and is larger when correlation is high.

Consider, however, the results when the experimenter uses correct prior knowledge about variance of the prior knowledge if that knowledge has a biased mean. Table 4.2 displays results extracted from both computer programs. Biases in the mean are given in terms of the prior standard deviation used, $\sqrt{d}$. An example will clarify usage of the table. If the experimenter uses the correct prior variance, d = 1, say, and his prior knowledge of the mean of this prior is biased by twice the prior standard deviation he uses, i.e., 2, then when $\rho = .85$, the true mean square error of the estimate is thirty-one percent of the "no prior" variance (mean square error) of $\hat{\beta}_1$. This may seem to imply that a bias of $2d^{1/2}$ in the mean of the prior is unimportant.

However, the true $MSE_w\tilde{\beta}_1$ is 1.81 times the mean square error indicated to the experimenter, i.e.,

$$MSE_w\tilde{\beta}_1 = 1.81V\tilde{\beta}_1.$$

One should note the following:

1. For d's of $100\sigma^2$ or more, biases of twice the standard deviation of the prior cause no bad effects, but the reduction in mean square error caused by using prior knowledge is not great, even when correlation is high.

2. For $d = \sigma^2$, when X'X is diagonal, a bias of the same size as the standard deviation of the prior does not inflate mean square error. Rather, $V\hat{\beta}_1$ is still reduced by one-fourth. However, the experimenter's available information states that the mean square error is actually two-thirds of its correct value. (1.50, in Table 4.2.) When the correlation is 0.96, the reduction is to six percent of $V\hat{\beta}_1$, and the experimenter's information is nearly correct. ($1.18V\tilde{\beta}_1$; the experimenter knows $V\tilde{\beta}_1$.)

None of the results shown in Table 4.2 imply that bias in the mean on the order of the size of the standard deviation of the prior used will adversely affect experimental conclusions unless the prior variance used is on the order of σ^2. However, bias larger than this may cause the experimenter to err, especially if he has not attached a high variance to his prior knowledge. Also, the errors will be smaller if correlation is high.

It is probable, however, that errors will be made in specifying both the mean and the variance of the prior. Table 4.3 displays results which show the effect of this type of error. Results are shown for both an over-estimation and an under-estimation of the prior variance by a five factor. Biases are expressed in terms of the prior standard deviations used.

There is no result shown in the above table, where the experimenter has overestimated prior variance ($d = 5d_0$), which implies that prior knowledge should not be used, unless biases in the mean exceed the standard deviation of the prior. If correlation is high, even biases in excess of the standard deviation of the prior do not have serious effects. Overestimation of prior variances would seem to be a safeguard against biases in the mean of the prior, to be used when the experimenter is not sure of the validity of his prior knowledge. The safeguard is most effective if X'X is ill-conditioned.

When the experimenter underestimates the variance of the prior, the opposite is the result. Table 4.3 ($d = d_0/5$) shows no case, excepting those with very low biases when d and d_0 are 100 or more, and when correlation is low, where errors introduced are tolerable. Either $MSE_w\tilde{\beta}_1$ is actually higher than $V\hat{\beta}_1$, or $MSE_w\tilde{\beta}_1$ is much larger than $V\tilde{\beta}_1$; often both conditions occur.

Added insight can be gained by considering, in Table 4.3, $d = 10$, $d_0 = 2$, and $d = .5$, $d_0 = 2.5$; the true prior variances (d_0) are nearly the same; the reduction from $V\hat{\beta}_1$ is large in each case. That is, $MSE_w\tilde{\beta}_1$ is between 13 and 16 percent of $V\hat{\beta}_1$ for $\rho = .96$. However, when the experimenter overestimates prior variance, he always believes that the mean square error is larger than it actually is; when he underestimates, he believes that it is much smaller than the true mean square error (as shown by the results .16;5.65 when correlation is .96 and bias is $2d^{1/2}$). This type of error has historically been considered undesirable.

The effect of assuming an incorrect prior variance but a correct mean is also shown in Table 4.3, where $\underline{m} = \underline{0}$. Again, overestimation of prior variance is more preferable than is underestimation.

Often, engineers prefer to think in terms of the weight given to prior knowledge. Section 5.2 develops an analogy to the above in terms of weights.

Table 4.4 was extracted from the simulation program described in Section 4.2. In that program, the mean square error of the estimators of β_1 and β_2 was calculated for every situation considered. Table 4.4 presents MSE$\underline{\beta}_H$, where $\underline{\beta}_H$ is the Hoerl estimator of $\underline{\beta}$, given in Chapter 3. This estimation technique can be expressed in terms of null prior knowledge. That is, if $\underline{\beta}_0$ is zero, and is given a variance of D, then Hoerl's situation is at hand, where, if D = dI, 1/d is added to the diagonal elements of X'X.

The purpose of Table 4.4 is to allow a judgment of the effect of levels of one parameter on the estimation of the other parameter when Hoerl's technique is used. In viewing the table, the following points should be considered.

1. The models are defined in column one, and range from $y = \varepsilon$ ($\beta_1,\beta_2 = 0.0$), to $y = 5x_1 + 5x_2 + \varepsilon$ ($\beta_1,\beta_2 = 5,5$).

2. The same generating seed was used to start the simulations for each model. For example, model 1 was chosen, a seed was picked, and

Table 4.4. $MSE\beta_{1_H}$; $MSE\beta_{2_H}$: the null prior and estimation of $\underline{\beta}$

β_1,β_2	Correlation 0	.50	.85	.96	d
$V\hat{\underline{\beta}}$	.93; 1.02	1.17; 1.48	3.83; 3.68	12.38;11.89	
0,0	.93; 1.02	1.17; 1.48	3.79; 3.63	11.78;11.30	
0,1	.93; 1.02	1.17; 1.48	3.79; 3.63	11.78;11.30	
0,5	.91; .89	1.07; 1.53	3.23; 2.47	8.17; 7.88	
1,0	.93; 1.02	1.17; 1.48	3.79; 3.63	11.78;11.30	
1,1	.93; 1.02	1.17; 1.48	3.79; 3.63	11.78;11.30	1000.0
1,5	.91; .89	1.07; 1.53	3.23; 2.47	8.17; 7.89	
5,0	.93; 1.02	1.17; 1.48	3.79; 3.63	11.78;11.31	
5,1	.93; 1.02	1.17; 1.48	3.79; 3.63	11.78;11.31	
5,5	.91; .89	1.07; 1.53	3.22; 2.47	8.18; 7.90	
0,0	.91; 1.00	1.13; 1.44	3.40; 3.25	7.91; 7.52	
0,1	.91; 1.00	1.14; 1.44	3.40; 3.25	7.92; 7.52	
0,5	.89; .87	1.04; 1.48	2.95; 2.24	5.68; 5.40	
1,0	.91; 1.00	1.13; 1.44	3.40; 3.26	7.92; 7.54	
1,1	.91; 1.00	1.13; 1.44	3.40; 3.25	7.91; 7.52	100.0
1,5	.89; .87	1.04; 1.48	2.94; 2.23	5.61; 5.33	
5,0	.91; 1.00	1.13; 1.43	3.43; 3.29	8.20; 7.83	
5,1	.91; 1.00	1.13; 1.43	3.42; 3.28	8.10; 7.73	
5,5	.89; .87	1.04; 1.48	2.91; 2.21	5.52; 5.28	
0,0	.77; .84	.87; 1.11	1.52; 1.43	1.19; 1.06	
0,1	.77; .85	.88; 1.13	1.55; 1.46	1.31; 1.18	
0,5	.75; .89	.89; 1.43	2.24; 2.21	3.70; 4.22	
1,0	.77; .84	.87; 1.11	1.57; 1.48	1.33; 1.20	
1,1	.77; .85	.86; 1.12	1.53; 1.43	1.19; 1.06	10.0
1,5	.75; .89	.83; 1.38	1.87; 1.85	2.55; 3.08	
5,0	.94; .84	1.13; 1.16	2.80; 2.29	4.86; 4.02	
5,1	.94; .85	1.08; 1.12	2.43; 1.92	3.68; 2.84	
5,5	.90; .89	.87; 1.22	1.40; 1.00	.94; .82	

t e n ε's were generated for each simulation on that model. Then model 2 was chosen, and the original seed was again used for each simulation on model 2.

3. Models with β_2 levels of zero and one were simulated 250 times; when β_2 was five, 50 simulations were performed. That is, the total set of random variables used in the two situations was not the same, even though the seeds were chosen as described in (2) above. Consequently, some results where $\beta_2 = 5$ are not realistic.

Table 4.4 shows the following:

1. When correlation is high, and a parameter is small, the level (magnitude) of the other parameter is directly associated with an increase in mean square error of the estimator of the small parameter.

2. If a large parameter is being estimated, the mean square error of its estimator decreases as the size of the other parameter increases.

3. The effects are evident when d is small and correlation is high; e.g., consider d = 10 and ρ = .96 in Table 4.4.

4.1.2. The Effect of Prior Information about $\underline{\beta}$ on Variance Estimation

The relationship of the variance of $\hat{\sigma}^2$ to correlation and the addition of correct prior knowledge is more stable than is the situation when $\hat{\underline{\beta}}$ is considered. $V\hat{\sigma}^2$ is reduced, from $2\sigma^4/(n-p)$ to $2\sigma^4/n$ (from .25 to .20 in the computations considered here).

In the following, the effect of errors in prior specification of the mean, of the variance, and of both mean and variance, are discussed.

Results pertaining to mean specification errors when variance is correctly specified are shown in Table 4.5, where ratios of mean square errors are given. Results when the bias is zero are also included, for comparison purposes. These results are from the ratios $V\tilde{\sigma}^2/V\hat{\sigma}^2$ and $V\tilde{\sigma}^2/V\tilde{\sigma}^2$ and are (0.80, 1.00) in all cases.

When biases are of the magnitude of the standard deviation of the prior variance, the true mean square error is greater than $V\hat{\sigma}^2$, and the disparity decreases as the prior variance used decreases. The variance used by the experimenter ($V\tilde{\sigma}^2$) is always smaller than the true mean square error. For example, when the prior mean is biased by two standard deviations and d is one, $\rho = .85$, $MSE_w\tilde{\sigma}^2$ is 2.71 times as great as is $V\hat{\sigma}^2$, and the ratio of $MSE_w\tilde{\sigma}^2$ to $V\tilde{\sigma}^2$ is 3.39. This means that the experimenter believes that his mean square error is less than one-third of its true value (i.e., less than $V\hat{\sigma}^2$). Also, one should note that the highly correlated cases show an increase in errors caused by bias, which is the reverse of the evidence drawn from viewing mean square error of the β_1 estimator.

Table 4.5 also shows that if biases are less than half of the standard deviation of the prior, the effect of bias is trivial.

When the mean is correctly specified, but the experimenter errs in his choice of prior variance, losses due to the use of priors are very likely unless he has overestimated his prior variance. Table 4.6 shows that for no case of overestimation ($d = 5d_0$), are there losses. For all cases of underestimation ($d = d_0/5$), excepting when prior variance is low ($d = 0.5$), losses are extremely large.

Table 4.5. $MSE_w\tilde{\sigma}^2/V\hat{\sigma}^2$; $MSE_w\tilde{\sigma}^2/V\tilde{\sigma}^2$; correctly specified prior variance

$md^{-1/2}$	Correlation 0	0.50	0.85	0.96	d
0.0	0.80;1.00	0.80;1.00	0.80;1.00	0.80;1.00	1000.0
0.5	0.89;1.11	0.89;1.11	0.89;1.11	0.89;1.11	1000.0
1.0	1.28;1.60	1.28;1.60	1.28;1.60	1.28;1.60	1000.0
2.0	4.63;5.79	4.64;5.80	4.64;5.80	4.64;5.80	1000.0
0.0	0.80;1.00	0.80;1.00	0.80;1.00	0.80;1.00	100.0
0.5	0.89;1.11	0.89;1.11	0.89;1.11	0.89;1.11	100.0
1.0	1.27;1.59	1.28;1.60	1.28;1.60	1.28;1.60	100.0
2.0	4.58;5.72	4.60;5.75	4.61;5.76	4.61;5.76	100.0
0.0	0.80;1.00	0.80;1.00	0.80;1.00	0.80;1.00	10.0
0.5	0.88;1.10	0.88;1.11	0.89;1.11	0.89;1.11	10.0
1.0	1.22;1.53	1.24;1.55	1.25;1.56	1.25;1.56	10.0
2.0	4.08;5.10	4.25;5.31	4.32;5.40	4.34;5.42	10.0
0.0	0.80;1.00	0.80;1.00	0.80;1.00	0.80;1.00	1.0
0.5	0.84;1.05	0.85;1.06	0.86;1.07	0.86;1.07	1.0
1.0	1.00;1.25	1.05;1.31	1.08;1.34	1.08;1.35	1.0
2.0	2.08;2.60	2.49;3.11	2.71;3.39	2.77;3.46	1.0
0.0	0.80;1.00	0.80;1.00	0.80;1.00	0.80;1.00	0.5
0.5	0.83;1.04	0.84;1.05	0.84;1.05	0.84;1.05	0.5
1.0	0.92;1.16	0.97;1.21	0.99;1.24	1.00;1.25	0.5
2.0	1.51;1.89	1.82;2.27	2.01;2.51	2.06;2.58	0.5

Consider the results for d = 0.5, when d_0 is 2.5 (lower table), with the results for d = 10, when d_0 = 2 (upper table). In these cases, the correct prior variances are nearly equal. For all correlations, $MSE_w\tilde{\sigma}^2$ is approximately 73 percent of $V\hat{\sigma}^2$ while $MSE_w\tilde{\sigma}^2$ is less than the variance the experimenter believes he has achieved. (92 percent of $V\tilde{\sigma}^2$). When underestimating, the experiment always achieves a larger $MSE_w\tilde{\sigma}^2$ than is $V\hat{\sigma}^2$, and the experimenter always believes that this variance is smaller than its actual value.

Table 4.6. $MSE_w\tilde{\sigma}^2/V\hat{\sigma}^2$; $MSE_w\hat{\sigma}^2/V\tilde{\sigma}^2$; incorrect prior variance specification on $\underline{\beta}$

Correlation 0	0.50	0.85	0.96	d	
0.75;0.94	0.75;0.94	0.74;0.94	0.75;0.93	1000.0	
0.75;0.93	0.75;0.93	0.74;0.93	0.74;0.92	100.0	
0.74;0.92	0.73;0.92	0.73;0.91	0.73;0.92	10.0	$d=5d_0$
0.72;0.90	0.73;0.91	0.74;0.92	0.75;0.93	1.0	
0.74;0.92	0.74;0.93	0.75;0.94	0.75;0.94	0.5	
7.19;8.99	7.19;8.98	7.16;8.95	7.05;8.82	1000.0	
7.09;8.86	7.05;8.81	6.82;8.52	6.06;7.58	100.0	
6.20;7.74	5.96;7.44	4.94;6.17	3.81;4.76	10.0	$d=d_0/5$
2.72;3.40	2.56;3.20	2.25;2.81	2.12;2.66	1.0	
1.80;2.24	1.74;2.18	1.65;2.06	1.61;2.01	0.5	

One should note also that the bad effect of overestimating decreases with d, the prior variance used. Table 4.6 shows clearly that the curve of "maximum benefit: and its dependence on n, ρ, d, and d_0 could be traced by a more detailed study.

The effect of incorrect variance specification on the estimation of σ^2 is more complex than can be determined by the cases examined in this thesis. Section 5.3 discusses a project currently being developed which will examine this problem more specifically, based on the general information gathered thus far.

When variance specification errors occur and bias errors are also made, it is seen, in Table 4.7, that unless bias is as small as the standard deviation of the prior, bias errors coupled with overestimation of prior variance do lead to large errors in judgment on the part of the experimenter. In the case of underestimation, added bias errors cause the experimenter to misjudge $MSE\tilde{\sigma}^2$ considerably at all times.

Table 4.7. $MSE_w \tilde{\sigma}^2/V\hat{\sigma}^2$; $MSE_w \tilde{\sigma}^2/V\tilde{\sigma}^2$; incorrect prior mean and variance specification on $\underline{\beta}$

$\underline{m}\ d^{-1/2}$	Correlation 0	.50	.85	.96	d
			$d = 5d_0$		
0.5	.71; .89	.71; .89	.72; .89	.72; .90	
1.0	.72; .90	.72; .90	.72; .90	.75; .93	1000.0
2.0	2.54; 3.17	2.54; 3.18	2.55; 3.19	2.66; 3.32	
0.5	.71; .89	.71; .89	.71; .89	.75; .94	
1.0	.72; .90	.72; .90	.74; .92	.91; 1.14	100.0
2.0	2.52; 3.15	2.54; 3.17	2.63; 3.28	3.34; 4.17	
0.5	.71; .89	.71; .89	.72; .91	.77; .97	
1.0	.73; .92	.74; .93	.82; 1.03	1.01; 1.26	10.0
2.0	2.31; 2.88	2.46; 3.07	2.83; 3.54	3.56; 4.45	
0.5	.72; .90	.72; .91	.74; .93	.75; .94	
1.0	.74; .92	.76; .95	.80; .99	.82; 1.02	1.0
2.0	1.25; 1.56	1.54; 1.93	1.78; 2.22	1.87; 2.33	
0.5	.74; .92	.74; .92	.75; .94	.76; .94	
1.0	.74; .93	.76; .95	.78; .97	.79; .98	0.5
2.0	.97; 1.21	1.16; 1.45	1.31; 1.63	1.36; 1.70	
			$d = d_0/5$		
0.5	7.92; 9.90	7.91; 9.89	7.89; 9.86	7.78; 9.73	
1.0	10.22;12.78	10.22;12.78	10.19;12.74	10.10;12.63	1000.0
2.0	21.24;26.55	21.24;26.56	21.22;26.52	21.16;26.45	
0.5	7.80; 9.75	7.77; 9.71	7.53; 9.42	6.80; 8.50	
1.0	10.07;12.59	10.05;12.56	9.81;12.26	9.12;11.41	100.0
2.0	20.90;26.13	20.93;26.16	20.68;25.85	20.21;25.26	
0.5	6.80; 8.51	6.59; 8.23	5.56; 6.95	4.42; 5.53	
1.0	8.73;10.91	8.59;10.74	7.54; 9.42	6.37; 7.96	10.0
2.0	17.92;22.40	18.18;22.72	17.08;21.35	15.79;19.74	
0.5	2.91; 3.63	2.80; 3.50	2.51; 3.14	2.39; 2.98	
1.0	3.50; 4.38	3.57; 4.47	3.34; 4.18	3.22; 4.03	1.0
2.0	6.32; 7.90	7.30; 9.13	7.43; 9.29	7.36; 9.20	
0.5	1.88; 2.35	1.86; 2.33	1.79; 2.23	1.76; 2.20	
1.0	2.14; 2.68	2.25; 2.81	2.23; 2.79	2.22; 2.77	0.5
2.0	3.39; 4.24	4.11; 5.14	4.44; 5.55	4.51; 5.63	

As was mentioned above, more detailed studies will be required before valid conclusions can be made about variance estimation. More models, and models estimated from field data, must be examined. The implications of these findings as they affect parameter ($\underline{\beta}$) estimation, and acceptance or rejection of hypotheses about β_1, are discussed in Chapter 5.

4.1.3. Variance of the Predicted Value

Table 4.9 shows the relationship between correlation and the effect of prior knowledge on the variance of the prediction. The points at which this variance was calculated are shown in Figure 4.1 as $p_1(-.05,0.5)$ and $p_2(0.4,-0.5)$. The relationships discussed are similar for both points; Tables 4.9 and 4.10 show ratios of mean squares, and reflect results for p_2 only. Although the point p_2 will not exemplify results when the experimenter extrapolates, it is outside the range of (x_1,x_2) for which the model has been estimated. That is, p_2 is out of the (two dimensional) range of X, although p_{21} and p_{22} are within the range of x_1 and x_2, respectively. Actual values of variance of the prediction are given in Table 4.8, for both points. The following results are indicated:

1. The decrease in the variance of the prediction caused by the addition of prior knowledge depends heavily on correlation, and on the variance of the prior knowledge. (Tables 4.8 and 4.9)

2. An error in the choice of this variance can be dangerous. As Table 4.10 shows, if the experimenter underestimates prior variance, the correct variance of the prediction is higher than that variance

($V\hat{y}_0$) available when prior information is not used. Errors are not large, however, until the prior variance used is as low as ten times the residual variance. At this level of d, or lower, $V_w\tilde{y}_0$ exceeds $V\hat{y}_0$ and/or $V\tilde{y}_0$.

3. Overestimation of prior variance does not cause the experimenter to make "bad" errors, if his belief that $V\tilde{y}_0$ is larger than its true value is not an error type which concerns him.

Comparison of the effects of over and under estimation can be seen in another light by considering cases in Table 4.10 where the true prior variances are nearly the same. Consider d = 10, when $d = 5d_0$, with d = 0.5 when $d = d_0/5$. Here the true prior variances are 2 and 2.5 respectively. When the experimenter overestimates, his correct variance ($V_w\tilde{y}_0$) is much smaller than $V\hat{y}_0$ and $V\tilde{y}_0$, and he believes the variance to be $V\tilde{y}_0$. When he underestimates prior variance, $V_w\tilde{y}_0$ is smaller than $V\hat{y}_0$ when correlation is high, but is somewhat larger than he believes it to be.

$MSE_w\tilde{y}_0$ is nearly the same as $V_w\tilde{y}_0$ for the cases studied in this paper. Biases of two (which range from six percent of the standard deviation of the prior to 241 percent of that standard deviation) cause maximum increments in $V_w\tilde{y}_0$ of 0.02. That is, $(\text{bias})^2 = 0.02$ for the calculation of $MSE_w\tilde{y}_0$ from $V_w\tilde{y}_0 + (\text{bias})^2$, for the case of maximum bias.

In summary, if the experimenter is conservative in his appraisal of the validity of his prior knowledge, and thus gives it high variance, then the variance of the predicted value will be decreased, and his results will err on the conservative side. He will obtain lower

Table 4.8. Variance of the prediction

Points	Correlation 0 p_1	0 p_2	.5 p_1	.5 p_2	.85 p_1	.85 p_2	.96 p_1	.96 p_2	d
				Correct Prior Variances					
$V(\hat{y}_0)$	.25	.41	.37	.81	1.06	2.71	3.99	10.57	
$V(\tilde{y}_0)$	.25	.41	.37	.81	1.06	2.70	3.90	10.30	1000.0
	.23	.37	.32	.68	.66	1.63	1.14	2.93	10.0
	.13	.21	.14	.27	.17	.35	.18	.39	1.0
	.08	.14	.09	.16	.10	.19	.10	.20	0.5
				Incorrect Prior Variances (overestimation: $d = 5d_0$)					
$V_w(\tilde{y}_0)$	.25	.41	.37	.81	1.06	2.68	3.82	10.09	1000.0
	.21	.35	.29	.59	.46	1.11	.50	1.23	10.0
	.08	.12	.07	.13	.07	.11	.06	.09	1.0
	.02	.04	.04	.06	.03	.05	.03	.04	0.5
				(underestimation: $d = d_0/5$)					
$V_w(\tilde{y}_0)$	.25	.41	.37	.82	1.09	2.77	4.29	11.35	1000.0
	.31	.51	.50	1.13	1.64	4.23	4.31	11.40	10.0
	.38	.62	.47	.99	.68	1.58	.78	1.90	1.0
	.31	.50	.35	.68	.41	.89	.44	.98	0.5

Table 4.9. $V\tilde{y}_0/V\hat{y}_0$

Correlation 0	0.5	0.85	0.96	d
1.00	1.00	1.00	0.97	1000.0
0.90	0.84	0.60	0.28	10.0
0.51	0.33	0.13	0.04	1.0
0.34	0.20	0.07	0.02	0.5

Table 4.10. $V_w\tilde{y}_0/V\hat{y}_0$; $V_w\tilde{y}_0/V\tilde{y}_0$

Correlation 0	0.5	0.85	0.96	d
		$d = 5d_0$		
1.00;1.00	1.00;1.00	0.99;0.99	0.95;0.98	1000.0
0.85;0.95	0.73;0.87	0.41;0.68	0.12;0.42	10.0
0.29;0.57	0.16;0.48	0.04;0.31	0.01;0.23	1.0
0.10;0.29	0.07;0.38	0.02;0.26	0.00;0.20	0.5
		$d = d_0/5$		
1.00;1.00	1.01;1.01	1.02;1.03	1.07;1.10	1000.0
1.24;1.38	1.40;1.66	1.56;2.60	1.08;3.89	10.0
1.51;2.95	1.22;3.67	0.58;4.51	0.18;4.87	1.0
1.22;3.57	0.84;4.25	0.33;4.68	0.09;4.90	0.5

variances than those available from using no prior knowledge, and those variances obtained will be higher than the true variances.

4.1.4. The Effect of Prior Knowledge on σ^2 on Variance Estimation

$MSE_w\tilde{\sigma}_c^2$, the mean square error of the estimator of σ^2 when incorrect prior knowledge has been used with models (2.1) and (2.2), is available from equations (2.24) and (2.25), and can be compared with related properties as follows:

1. Comparison with equation (2.16) will relate the use of incorrect prior knowledge on σ^2 to properties of the "no prior" estimator.

2. Comparison with equation (2.19) will give the relationship with the use of correct prior knowledge on $\underline{\beta}$.

3. Comparison with equation (2.20) will show the relationships with the use of correct prior knowledge on σ^2 and $\underline{\beta}$.

4. Comparison with equation (2.23) will allow judgments to be made about how the use of incorrect prior knowledge on $\underline{\beta}$ compares with the use of incorrect prior knowledge on σ^2 as they both affect the estimator of σ^2.

Even in the simple two parameter case, mathematical relationships which will show when prior knowledge on σ^2 should be used are quite complex, algebraically. The ideas which these equations display are only clear after the application of restrictive assumptions. For the cases examined in this thesis, the pertinent quantities were calculated, and are displayed in Tables 4.11 - 4.14, as ratios of mean square errors.

Relationships concerning MSE $_w\tilde{\sigma}_c^2$ are more complex than are those pertaining to $\text{MSE}_w\tilde{\sigma}^2$ and $\text{MSE}_w\underline{\tilde{\beta}}$. Another study could be undertaken, examining the estimation of σ^2 specifically, using findings given here as a guide.

The effect of adding correct prior knowledge can be seen by noting that

$$V\hat{\sigma}^2 = 2\sigma^4/(n-p);\ V\tilde{\sigma}^2 = 2\sigma^4/n;\ V\tilde{\sigma}_c = 2\sigma^4/(n+k)\ .$$

Thus, the pertinent ratios $V\tilde{\sigma}_c^2/V\hat{\sigma}^2$ and $V\tilde{\sigma}_c^2/V\tilde{\sigma}^2$ are $(n-p)/(n+k)$ and $n/(n+k)$. If the degrees of freedom are correct, and the prior distribution of c/σ^2 is X_k^2, then advantage is always gained and correlation plays no role.

When the experimenter has used prior knowledge on $\underline{\beta}$ with no bias, and has added correct knowledge on σ^2, but has erred in his choice of

Table 4.11. $MSE_w \tilde{\sigma}^2_c / V\tilde{\sigma}^2$; $MSE_w \tilde{\sigma}^2_c / V\tilde{\sigma}^2_c$; mean, degrees of freedom, and prior correct, and variance specification incorrect

k	Correlation 0	.50	.85	.96	d
			$d = 5d_0$ (Overestimation)		
8	.54; .96 }	These figures are approximately the same for all prior variances (d), and for all correlations (that is, they are correct within .01)			
16	.38; .98 }				
32	.23; .99 }				
			$d = d_0/5$ (Underestimation)		
8	3.02;5.44	3.02;5.43	3.01;5.42	2.97;5.34	
16	1.57;4.07	1.57;4.07	1.56;4.06	1.54;4.01	1000.0
32	.69;2.90	.69;2.90	.69;2.89	.68;2.86	
8	2.98;5.37	2.97;5.34	2.88;5.18	2.57;4.65	
16	1.55;4.02	1.54;4.01	1.50;3.89	1.36;3.53	100.0
32	.68;2.87	.68;2.86	.66;2.79	.61;2.57	
8	2.63;4.75	2.54;4.58	2.15;3.87	1.72;3.09	
16	1.38;3.59	1.34;3.48	1.15;2.99	.94;2.45	10.0
32	.62;2.61	.60;2.53	.53;2.23	.45;1.90	
8	1.30;2.33	1.23;2.22	1.11;2.01	1.06;1.92	
16	.74;1.92	.71;1.85	.65;1.70	.63;1.64	1.0
32	.37;1.57	.36;1.52	.34;1.43	.33;1.39	
8	.94;1.69	.92;1.65	.88;1.59	.87;1.56	
16	.57;1.48	.56;1.45	.54;1.41	.54;1.39	0.5
32	.31;1.30	.31;1.28	.30;1.25	.30;1.24	

the variance of the prior knowledge on $\underline{\beta}$, the following results are indicated, from Table 4.11.

1. Correlation is not an important factor.

2. The more information attached to σ^2 (high degrees of freedom) the less important are errors in variance specification. Given a specific prior variance (on $\underline{\beta}$), there is a level of degrees of freedom

for the prior on σ^2 which will dispel the bad effect of underestimation of the prior variance on $\underline{\beta}$.

3. Overestimation of prior variance is an error type that does not lead to bad results. As an example of the effect of underestimation, when ρ = .96, d = 1000, and k = 8, the correct mean square error, $MSE_w\tilde{\sigma}_c^2$, is 2.97 times as large as $V\tilde{\sigma}^2$, and 5.34 times as large as $V\tilde{\sigma}_c^2$, which is the information available to the experimenter. The ratio of $MSE_w\tilde{\sigma}_c^2$ to $V\hat{\sigma}^2$ is always as follows:

$$MSE_w\tilde{\sigma}_c^2/V\hat{\sigma}^2 = .8MSE_w\tilde{\sigma}_c^2/V\tilde{\sigma}^2 \ .$$

If the experimenter errs in prior specification, and if the true prior on σ^2 is exponential, and if he also makes variance specification errors, his results are as displayed in Table 4.12. Table 4.11, where the prior specification is correct, should be compared with Table 4.12. Again, correlation is not important, but prior specification error results in large increases in mean square error, especially for large k. The results are somewhat worse for underestimation than for overestimation.

Table 4.13 gives results when bias errors, on $\underline{\beta}$ prior knowledge, are null, and the prior on σ^2 is correct, i.e., χ^2 distributed. The effect of making errors in choosing degrees of freedom for the prior on σ^2 can be seen, when the prior variances attached to information on $\underline{\beta}$ are also wrong. Results when degrees of freedom, k, are correct are given in Table 4.11.

Table 4.12. $MSE_w\tilde{\sigma}_c^2/V\tilde{\sigma}^2$; $MSE_w\tilde{\sigma}_c^2/V\tilde{\sigma}_c^2$; mean, degrees of freedom correct, and prior and variance specifications incorrect; the true prior is exponential

k	Correlation 0	.50	.85	.96	d
		$d = 5d_0$			
8	1.27; 2.29 }	These figures are correct within 0.01, and			
16	2.03; 5.28 }	are approximately the same for all prior			
32	2.95;12.41 }	variances, d, and for all correlations.			
		$d = d_0$			
8	1.30; 2.33 }	Accurate within 0.01, these figures are			
16	2.04; 5.31 }	approximately the same for all prior variances,			
32	2.96;12.43 }	d, and for all correlations.			
		$d = d_0/5$			
8	3.76; 6.77	3.76; 6.77	3.75; 6.75	3.71; 6.68	
16	3.22; 8.38	3.22; 8.38	3.22; 8.37	3.20; 8.31	1000.0
32	3.41;14.33	3.41;14.33	3.41;14.32	3.40;14.29	
8	3.72; 6.70	3.71; 6.67	3.62; 6.51	3.32; 5.99	
16	3.20; 8.33	3.20; 8.31	3.15; 8.20	3.01; 7.84	100.0
32	3.41;14.30	3.40;14.29	3.39;14.22	3.33;14.00	
8	3.38; 6.08	3.29; 5.91	2.89; 5.21	2.46; 4.42	
16	3.04; 7.90	3.00; 7.79	2.80; 7.27	2.60; 6.75	10.0
32	3.34;14.03	3.32;13.96	3.25;13.66	3.17;13.32	
8	2.04; 3.67	1.98; 3.55	1.86; 3.34	1.81; 3.25	
16	2.40; 6.23	3.37; 6.15	2.31; 6.00	2.29; 5.94	1.0
32	3.10;13.00	3.08;12.95	3.06;12.86	3.05;12.82	
8	1.68; 3.03	1.66; 2.99	1.62; 2.92	1.61; 2.90	
16	2.23; 5.79	2.22; 5.76	2.20; 5.72	2.19; 5.70	0.5
32	3.03;12.73	3.03;12.71	3.02;12.68	3.02;12.67	

Table 4.13. $MSE_w\tilde{\sigma}_c/V\tilde{\sigma}^2$; $MSE_w\tilde{\sigma}_c^2/V\tilde{\sigma}_c^2$; mean and prior correct, degrees of freedom and variance specifications incorrect

k	k_0	Correlation 0	.50	.85	.96	d	d_0
8	4	0.86;1.54					
	12	0.71;1.28					
16	12	0.53;1.38	(.01)			1000.0	d/5
	20	0.46;1.19					
32	28	0.29;1.23					
	36	0.27;1.12					
8	4	0.80;1.45		These results are correct within the error noted in in parentheses, and are approximately the same for all correlations.			
	12	0.76;1.35					
16	12	0.50;1.31	(.06)			10.0	d/5
	20	0.47;1.24					
32	28	0.28;1.19					
	36	0.27;1.15					
8	4	0.71;1.29					
	12	0.85;1.53					
16	12	0.46;1.20	(.01)			0.5	d/5
	20	0.52;1.36					
32	28	0.27;1.12					
	36	0.29;1.23					
8	4	2.16;3.88	2.16;3.88	2.15;3.87	2.11;3.81		
	12	4.38;7.88	4.38;7.88	4.36;7.85	4.31;7.76		
16	12	1.15;3.00	1.15;2.99	1.15;2.99	1.13;2.95	1000.0	5d
	20	2.22;5.76	2.22;5.76	2.21;5.75	2.18;5.68		
32	28	0.53;2.24	0.53;2.24	0.53;2.23	0.53;2.20		
	36	0.94;3.95	0.94;3.95	0.94;3.93	0.93;3.90		
8	4	1.86;3.35	1.79;3.23	1.51;2.72	1.23;2.22		
	12	3.91;7.03	3.79;6.82	3.29;5.91	2.69;4.85		
16	12	1.01;2.63	0.98;2.54	0.84;2.19	0.71;1.85	10.0	5d
	20	1.99;5.17	1.93;5.03	1.69;4.40	1.41;3.66		
32	28	0.48;2.01	0.47;1.96	0.41;1.74	0.36;1.52		
	36	0.85;3.58	0.83;3.49	0.74;3.11	0.63;2.65		
8	4	0.73;1.32	0.73;1.32	0.73;1.32	0.74;1.33		
	12	1.64;2.95	1.60;2.88	1.53;2.74	1.49;2.69		
16	12	0.47;1.22	0.47;1.22	0.47;1.22	0.47;1.23	0.5	5d
	20	0.90;2.35	0.89;2.30	0.85;2.21	0.83;2.17		
32	28	0.27;1.14	0.27;1.14	0.27;1.14	0.27;1.14		
	32	0.44;1.84	0.43;1.81	0.42;1.75	0.41;1.72		

1. Overestimation of prior variance ($d_0 = d/5$)

a. The choice of degrees of freedom is not extremely important, but such errors increase mean square error whether the experimenter chooses too large <u>or</u> too small a value for k. Choosing k too small causes slightly larger errors. (Tables 4.11 and 4.13)

b. Overestimation of prior variance is to be preferred over underestimation; however, an error in the choice of k leads the experimenter to believe that the variance of the estimator of σ^2 is smaller than it really is.

2. Underestimation of prior variance on $\underline{\beta}$ ($d_0 = 5d$)

a. Choosing k smaller than its true value is to be preferred over the other alternative.

b. More information on σ^2 (large degrees of freedom) nullifies the bad effect of underestimation.

c. The bad effect of underestimation decreases as the prior variance used, d, decreases.

These results cannot be generalized, considering the ranges of parameters studied here. Rather, they should be used as a basis for further study.

Finally, when the prior on σ^2 has been correctly chosen, and errors are made in the choice of the prior on $\underline{\beta}$, results given in Table 4.14 indicate that

1. Biases in the mean of the prior cause large errors unless these biases are small relative to the prior standard deviation, and the errors caused are larger when the experimenter underestimates prior variance than when he overestimates that variance.

Table 4.14. $MSE_w\tilde{\sigma}_c^2/V\tilde{\sigma}^2$; $MSE_w\tilde{\sigma}_c^2/V\tilde{\sigma}_c^2$; degrees of freedom and prior correct, mean and variance specifications incorrect

k	Correlation 0	.50	.85	.96	$m/\sqrt{d}$	d	d_0
	0.52; 0.94	0.52; 0.94	0.52; 0.94	0.54; 0.96	0.5	100	20
8	0.52; 0.94	0.52; 0.94	0.53; 0.96	0.60; 1.08	1.0		
	1.22; 2.19	1.23; 2.21	1.26; 2.27	1.54; 2.76	2.0		
	0.23; 0.97	0.23; 0.97	0.23; 0.97	0.24; 0.99	0.5		
32	0.23; 0.98	0.25; 0.98	0.23; 0.98	0.25; 1.03	1.0		
	0.36; 1.51	0.36; 1.52	0.37; 1.54	0.42; 1.76	2.0		
	0.52; 0.94	0.53; 0.94	0.53; 0.96	0.54; 0.97	0.5	1	.2
8	0.53; 0.96	0.54; 0.97	0.55; 1.00	0.56; 1.01	1.0		
	0.73; 1.31	0.84; 1.52	0.93; 1.68	0.97; 1.74	2.0		
	0.23; 0.98	0.23; 0.98	0.23; 0.98	0.24; 0.99	0.5		
32	0.23; 0.98	0.24; 0.99	0.24; 1.00	0.24; 1.00	1.0		
	0.27; 1.13	0.29; 1.22	0.31; 1.29	0.31; 1.32	2.0		
	0.59; 1.06	0.59; 1.06	0.59; 1.06	0.61; 1.09	0.5	100	100
8	0.74; 1.33	0.74; 1.33	0.75; 1.34	0.81; 1.46	1.0		
	2.01; 3.62	2.02; 3.64	2.05; 3.69	2.30; 4.14	2.0		
	0.24; 1.02	0.24; 1.03	0.25; 1.03	0.25; 1.04	0.5		
32	0.27; 1.14	0.27; 1.14	0.27; 1.15	0.29; 1.19	1.0		
	0.51; 2.12	0.51; 2.13	0.51; 2.15	0.56; 2.35	2.0		
	0.57; 1.02	0.57; 1.03	0.57; 1.03	0.57; 1.03	0.5	1	1
8	0.61; 1.10	0.63; 1.14	0.64; 1.16	0.65; 1.17	1.0		
	0.96; 1.72	1.12; 2.02	1.22; 2.20	1.25; 2.25	2.0		
	0.24; 1.01	0.24; 1.01	0.24; 1.01	0.24; 1.01	0.5		
32	0.25; 1.04	0.25; 1.06	0.25; 1.07	0.26; 1.07	1.0		
	0.31; 1.31	0.34; 1.44	0.36; 1.51	0.37; 1.54	2.0		
	3.26; 5.86	3.24; 5.84	3.15; 5.68	2.87; 5.17	0.5	100	500
8	4.13; 7.43	4.12; 7.42	4.03; 7.25	3.77; 6.78	1.0		
	8.31;14.96	8.32;14.98	8.22;14.80	8.05;14.48	2.0		
	0.73; 3.08	0.73; 3.07	0.72; 3.00	0.66; 2.79	0.5		
32	0.90; 3.76	0.89; 3.75	0.88; 3.68	0.83; 3.48	1.0		
	1.66; 6.98	1.66; 6.99	1.65; 6.92	1.61; 6.78	2.0		

Table continued

Table 4.14. (continued)

k	Correlation 0	.50	.85	.96	$m/\sqrt{d}$	d	d_0
	1.37; 2.46	1.33; 2.39	1.22; 2.19	1.17; 2.10	0.5		
8	1.60; 2.88	1.63; 2.93	1.54; 2.76	1.49; 2.68	1.0		
	2.69; 4.83	3.07; 5.52	3.11; 5.60	3.09; 5.56	2.0		
						1	5
	0.39; 1.63	0.38; 1.60	0.36; 1.51	0.35; 1.47	0.5		
32	0.43; 1.80	0.44; 1.83	0.42; 1.76	0.41; 1.72	1.0		
	0.63; 2.64	0.70; 2.94	0.71; 2.97	0.70; 2.95	2.0		

2. These biases in the mean are not effective until the mean is on the order of magnitude of the standard deviation of the prior.

3. Increasing the information on σ^2 (applying large degrees of freedom, k) decreases these bad effects.

4.2. Power of Tests of Hypotheses and Prior Knowledge

Figure 4.3 outlines the simulation program used to

1. Examine the power curve of the t-test as it changes with the condition of X'X and the addition of prior knowledge.

2. Examine the power curve when the prior knowledge is incorrect.

3. Consider the estimator used by Hoerl, analogous, in the calculation of $\underline{\beta}$ estimates, to the addition of null prior knowledge.

All combinations of possibilities of incorrect priors, models, and variances used by the experimenter were calculated by the program in one simulation; two hundred fifty simulations were performed. For the consideration of priors biased in their means, one hundred simulations were performed. A count was made of the number of excesses of

<u>Models</u>:

$y = 0$; $y = x_2$

$y = x_1$; $y = x_1+x_2$

$y = 5x_1$; $y = 5x_1+x_2$

$\rho = (0;0.5;0.85;0.96)$

$d = (100;100;10)$; $D = dI$

$D_0 = (5D;D;D/5)$; $D_0 = d_0 I$

$\underline{m}I_p = (0_{pp};D^{+1/2};2D^{+1/2})$

<u>True Priors</u>: Normal
Beta
Uniform
Exponential

H_0: $\beta_1=0$; H_a: $\beta_1 = \beta_1^*$

Level: 5 per cent

<u>Statistic</u>:

$$\frac{\underline{\bar{\beta}}'X^{*\prime}\underline{y}^* - \bar{\beta}_2\underline{x}_2^{*\prime}\underline{y}^*}{(\underline{y}^{*\prime}\underline{y}^*-\underline{\bar{\beta}}'X^{*\prime}\underline{y}^*)/(\text{d.f.})}$$

Test 1: $\underline{\bar{\beta}} = S^{-1}X^{*\prime}\underline{y}^*$; $\bar{\beta}_2 = \underline{x}_2^{*\prime}\underline{y}^*/\underline{x}_2^{*\prime}\underline{x}_2^*$

$\underline{y}^{*\prime} = (\underline{y}';\underline{0}')$, $X^{*\prime} = (X';0)$

Test 2: $\underline{\bar{\beta}} = \tilde{S}^{-1}X^{*\prime}\underline{y}^*$; $\bar{\beta}_2 = \underline{x}_2^{*\prime}\underline{y}^*/\underline{x}_2^{*\prime}\underline{x}_2^*$

$\underline{y}^{*\prime}=(\underline{y}';\underline{0}')$, $X^{*\prime}=(X';D^{-1/2})$

Test 3: $\underline{\bar{\beta}}$ and $\bar{\beta}_2$ are as in Test 2;

$\underline{y}^{*\prime}=(\underline{y}';\underline{\beta}_0'D^{-1/2})$, $X^{*\prime}=(X';D^{-1/2})$

Observations: 10

Program Definition

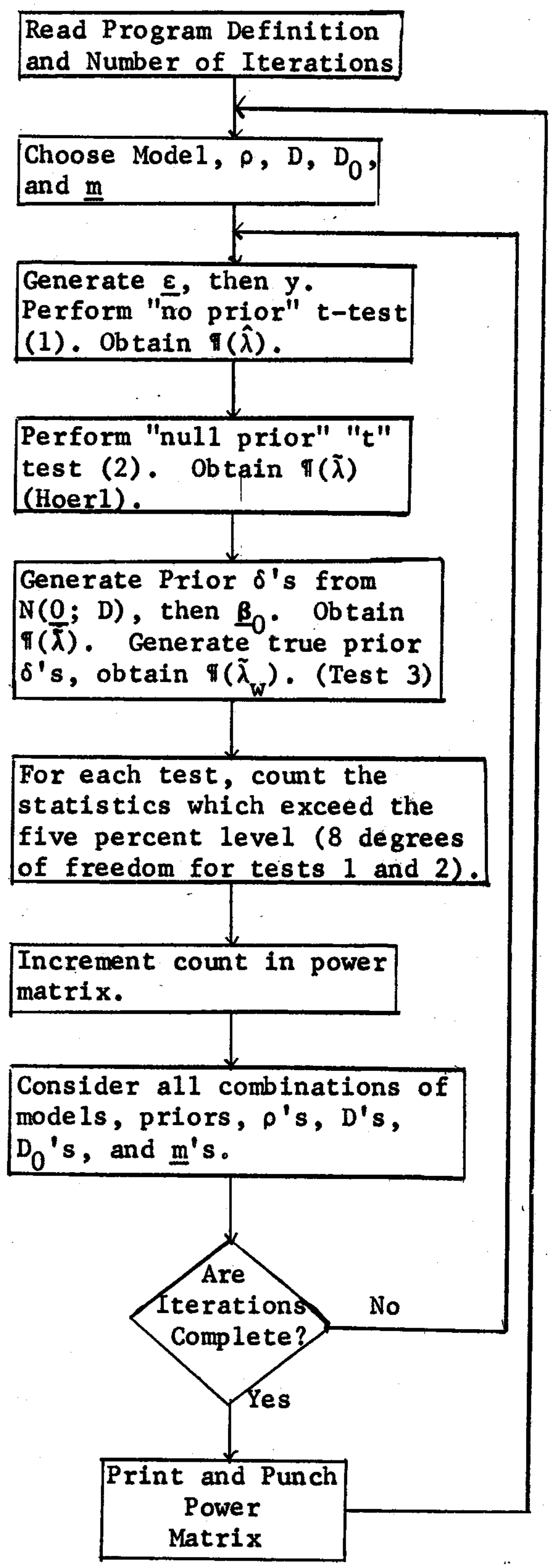

Figure 4.3. The program for the simulation of power

the statistics (3.16) and (3.17) over the five percent t value; the estimated power is the percentage of such excesses.

X matrices were chosen for the simulations so that correlations would be 0, .50, .85, and .96. (Actual correlations to four places were .0000, .4997, .8507, and .9617.)

The vector $\underline{\beta}_0$ was generated such that each element was from the same distribution (3.15). That is, if β_{0_1} was considered to be exponentially distributed, then β_{0_2} was distributed exponentially also.

Tables 4.15-4.20 were prepared by a computer. It is hoped that no confusion will result from the lack of lower case letters in computer output. In some cases in the tables D is a matrix. In other cases, e.g., 1/SQRT(D), D is the lower case d from D = dI.

The random number generators are discussed in Appendix 7.2, where the programs used for deviate generation are listed.

In the following, the important relationships between the ordinates of the power curve, the condition of X'X (i.e., the correlation), and the addition of incorrect prior knowledge are summarized.

In the consideration of power, much thought was given to deciding whether the size of the test should be kept constant by adjusting t_α^2, or whether the modifications to the total power curve should be viewed. The latter is more practically important to the experimenter, and in addition, in a simulation program, the adjustment of t_α^2 to a number which must be set approximately after many simulations requires much time and memory capability in the computer. Therefore it was decided to set t_α^2 and view the power curve over a range of the noncentrality parameter from zero to some positive number.

TABLE 4.15. POWER OF THE 'T' TEST FOR H(0), B(1) = 0, AS IT RELATES TO CORRELATION AND PRIOR KNOWLEDGE (OVERESTIMATION OF THE TRUE PRIOR VARIANCE)

D, THE PRIOR VARIANCE USED, IS 1000

DISTRIBUTION	CORRECT PRIORS[1] MEAN	VARIANCE	RHO = 0.00 B(1)=0,	1,	5	RHO = 0.50 B(1)=0,	1,	5	RHO = 0.85 B(1)=0,	1,	5	RHO = 0.96 B(1)=0,	1,	5
NORMAL	0	0.2*D	0.08	0.20	1.00	0.06	0.20	0.99	0.09	0.10	0.74	0.05	0.08	0.28
	1/SQRT(D)	0.2*D	0.06	0.14	1.00	0.03	0.07	0.98	0.07	0.08	0.56	0.03	0.05	0.17
	2/SQRT(D)	0.2*D	0.0	0.04	0.99	0.0	0.03	0.94	0.01	0.04	0.37	0.01	0.01	0.06
NORMAL	0	1.0*D	0.04	0.16	1.00	0.04	0.14	0.98	0.07	0.10	0.66	0.04	0.05	0.24
	1/SQRT(D)	1.0*D	0.05	0.11	1.00	0.01	0.09	0.98	0.03	0.08	0.52	0.01	0.02	0.16
	2/SQRT(D)	1.0*D	0.02	0.07	0.96	0.01	0.05	0.87	0.02	0.04	0.36	0.02	0.01	0.12
BETA	0	0.2*D	0.08	0.21	1.00	0.06	0.19	0.98	0.09	0.12	0.74	0.05	0.08	0.30
	1/SQRT(D)	0.2*D	0.06	0.15	1.00	0.01	0.10	0.98	0.05	0.09	0.54	0.02	0.02	0.15
	2/SQRT(D)	0.2*D	0.0	0.04	1.00	0.0	0.03	0.94	0.01	0.05	0.36	0.01	0.01	0.05
BETA	0	1.0*D	0.05	0.17	1.00	0.03	0.14	0.98	0.06	0.08	0.66	0.05	0.06	0.24
	1/SQRT(D)	1.0*D	0.02	0.12	1.00	0.02	0.08	0.98	0.07	0.08	0.48	0.02	0.02	0.15
	2/SQRT(D)	1.0*D	0.0	0.07	0.98	0.0	0.01	0.90	0.02	0.03	0.37	0.01	0.01	0.08
UNIFORM	0	0.2*D	0.07	0.21	1.00	0.06	0.18	0.98	0.09	0.12	0.74	0.05	0.08	0.29
	1/SQRT(D)	0.2*D	0.05	0.15	1.00	0.03	0.11	0.98	0.09	0.09	0.57	0.02	0.03	0.18
	2/SQRT(D)	0.2*D	0.0	0.05	0.98	0.0	0.02	0.93	0.01	0.03	0.40	0.01	0.0	0.06
UNIFORM	0	1.0*D	0.04	0.15	1.00	0.03	0.12	0.98	0.06	0.10	0.63	0.04	0.05	0.24
	1/SQRT(D)	1.0*D	0.02	0.11	0.99	0.0	0.05	0.96	0.04	0.06	0.51	0.01	0.03	0.13
	2/SQRT(D)	1.0*D	0.01	0.06	0.96	0.0	0.02	0.88	0.01	0.06	0.37	0.01	0.0	0.08
EXPONENTIAL	0	0.2*D	0.08	0.21	1.00	0.06	0.19	0.99	0.10	0.10	0.74	0.05	0.08	0.29
	1/SQRT(D)	0.2*D	0.04	0.11	1.00	0.01	0.11	0.98	0.06	0.08	0.55	0.02	0.02	0.18
	2/SQRT(D)	0.2*D	0.0	0.06	0.98	0.0	0.03	0.91	0.01	0.04	0.39	0.01	0.0	0.06
EXPONENTIAL	0	1.0*D	0.05	0.16	1.00	0.05	0.15	0.98	0.08	0.09	0.67	0.04	0.06	0.23
	1/SQRT(D)	1.0*D	0.05	0.15	0.99	0.02	0.13	0.97	0.07	0.08	0.58	0.01	0.02	0.13
	2/SQRT(D)	1.0*D	0.0	0.06	0.98	0.0	0.04	0.91	0.02	0.05	0.37	0.0	0.0	0.08
POWER USING NO PRIOR INFORMATION			0.05	0.15	1.00	0.04	0.12	0.98	0.07	0.09	0.64	0.04	0.05	0.21
POWER USING NULL PRIOR INFORMATION[2]			0.05	0.15	1.00	0.04	0.11	0.98	0.07	0.08	0.64	0.03	0.05	0.21

[1] THE EXPERIMENTER ADDS PRIOR KNOWLEDGE DISTRIBUTED N(0,D). POWER IS GIVEN FOR PRIORS DISTRIBUTED AS ABOVE.

[2] IN CALCULATING THE ESTIMATOR OF B, USING NULL PRIOR INFORMATION IS EQUIVALENT TO ADDING THE INVERSE OF D TO X'X.

TABLE 4.16. POWER OF THE 'T' TEST FOR H(0), B(1) = 0, AS IT RELATES TO CORRELATION AND PRIOR KNOWLEDGE (OVERESTIMATION OF THE TRUE PRIOR VARIANCE)

D, THE PRIOR VARIANCE USED, IS 100

DISTRIBUTION	CORRECT PRIORS[1] MEAN	VARIANCE	RHO = 0.00 B(1)=0,	1,	5	RHO = 0.50 B(1)=0,	1,	5	RHO = 0.85 B(1)=0,	1,	5	RHO = 0.96 B(1)=0,	1,	5
NORMAL	0	0.2*D	0.07	0.21	1.00	0.05	0.19	0.99	0.08	0.11	0.77	0.04	0.07	0.32
	1/SQRT(D)	0.2*D	0.06	0.15	1.00	0.02	0.11	0.99	0.07	0.09	0.54	0.03	0.02	0.24
	2/SQRT(D)	0.2*D	0.0	0.06	0.99	0.0	0.03	0.92	0.02	0.03	0.39	0.0	0.0	0.06
NORMAL	0	1.0*D	0.06	0.17	1.00	0.04	0.14	0.98	0.06	0.08	0.69	0.05	0.07	0.29
	1/SQRT(D)	1.0*D	0.03	0.15	1.00	0.01	0.08	0.98	0.07	0.08	0.56	0.02	0.02	0.21
	2/SQRT(D)	1.0*D	0.02	0.06	0.99	0.0	0.03	0.90	0.0	0.04	0.38	0.0	0.0	0.07
BETA	0	0.2*D	0.07	0.20	1.00	0.06	0.18	0.99	0.08	0.10	0.77	0.05	0.06	0.31
	1/SQRT(D)	0.2*D	0.03	0.16	1.00	0.03	0.12	0.98	0.05	0.09	0.61	0.03	0.02	0.23
	2/SQRT(D)	0.2*D	0.0	0.08	0.99	0.0	0.04	0.97	0.01	0.04	0.43	0.0	0.0	0.08
BETA	0	1.0*D	0.05	0.16	1.00	0.04	0.12	0.98	0.07	0.08	0.68	0.05	0.06	0.28
	1/SQRT(D)	1.0*D	0.06	0.13	1.00	0.01	0.11	0.98	0.05	0.08	0.58	0.01	0.01	0.22
	2/SQRT(D)	1.0*D	0.01	0.07	0.97	0.0	0.05	0.92	0.01	0.06	0.41	0.0	0.01	0.10
UNIFORM	0	0.2*D	0.08	0.19	1.00	0.05	0.20	0.99	0.10	0.12	0.76	0.05	0.07	0.33
	1/SQRT(D)	0.2*D	0.02	0.14	1.00	0.01	0.13	0.98	0.06	0.09	0.61	0.02	0.02	0.18
	2/SQRT(D)	0.2*D	0.01	0.09	0.97	0.0	0.03	0.92	0.0	0.05	0.38	0.0	0.0	0.08
UNIFORM	0	1.0*D	0.04	0.16	1.00	0.02	0.14	0.99	0.06	0.09	0.70	0.04	0.04	0.30
	1/SQRT(D)	1.0*D	0.02	0.14	1.00	0.01	0.14	0.97	0.07	0.07	0.53	0.04	0.03	0.21
	2/SQRT(D)	1.0*D	0.01	0.09	0.96	0.01	0.03	0.88	0.04	0.06	0.42	0.01	0.01	0.10
EXPONENTIAL	0	0.2*D	0.07	0.20	1.00	0.06	0.19	0.99	0.09	0.10	0.76	0.04	0.05	0.32
	1/SQRT(D)	0.2*D	0.04	0.14	1.00	0.01	0.12	0.98	0.06	0.08	0.59	0.01	0.01	0.23
	2/SQRT(D)	0.2*D	0.01	0.08	0.99	0.0	0.03	0.98	0.02	0.05	0.44	0.01	0.0	0.07
EXPONENTIAL	0	1.0*D	0.06	0.17	1.00	0.04	0.14	0.98	0.08	0.10	0.70	0.05	0.05	0.29
	1/SQRT(D)	1.0*D	0.05	0.14	0.99	0.02	0.11	0.97	0.05	0.10	0.60	0.03	0.02	0.22
	2/SQRT(D)	1.0*D	0.01	0.09	0.97	0.01	0.04	0.91	0.01	0.05	0.37	0.01	0.0	0.13
POWER USING NO PRIOR INFORMATION			0.05	0.15	1.00	0.04	0.12	0.98	0.07	0.09	0.64	0.04	0.05	0.21
POWER USING NULL PRIOR INFORMATION[2]			0.04	0.14	1.00	0.04	0.10	0.98	0.05	0.07	0.59	0.02	0.02	0.16

---

1 THE EXPERIMENTER ADDS PRIOR KNOWLEDGE DISTRIBUTED N(0,D). POWER IS GIVEN FOR PRIORS DISTRIBUTED AS ABOVE.

2 IN CALCULATING THE ESTIMATOR OF B, USING NULL PRIOR INFORMATION IS EQUIVALENT TO ADDING THE INVERSE OF D TO X'X.

TABLE 4.17. POWER OF THE 'T' TEST FOR H(0), B(1) = 0, AS IT RELATES TO CORRELATION AND PRIOR KNOWLEDGE (OVERESTIMATION OF THE TRUE PRIOR VARIANCE)

D, THE PRIOR VARIANCE USED, IS 10

DISTRIBUTION	CORRECT PRIORS[1] MEAN	VARIANCE	RHO = 0.00 B(1)=0,	1,	5	RHO = 0.50 B(1)=0,	1,	5	RHO = 0.85 B(1)=0,	1,	5	RHO = 0.96 B(1)=0,	1,	5
NORMAL	0	0.2*D	0.05	0.19	1.00	0.04	0.17	1.00	0.04	0.09	0.90	0.02	0.02	0.77
	1/SQRT(D)	0.2*D	0.06	0.20	1.00	0.01	0.17	0.99	0.03	0.09	0.85	0.01	0.02	0.59
	2/SQRT(D)	0.2*D	0.04	0.17	1.00	0.02	0.08	0.99	0.0	0.03	0.70	0.0	0.0	0.34
NORMAL	0	1.0*D	0.04	0.17	1.00	0.02	0.16	1.00	0.05	0.09	0.87	0.04	0.06	0.62
	1/SQRT(D)	1.0*D	0.04	0.23	1.00	0.03	0.11	1.00	0.05	0.05	0.76	0.04	0.06	0.49
	2/SQRT(D)	1.0*D	0.03	0.17	1.00	0.02	0.06	0.98	0.01	0.04	0.60	0.0	0.01	0.31
BETA	0	0.2*D	0.07	0.19	1.00	0.05	0.17	1.00	0.04	0.10	0.90	0.00	0.02	0.69
	1/SQRT(D)	0.2*D	0.05	0.21	1.00	0.02	0.16	1.00	0.03	0.09	0.85	0.01	0.02	0.61
	2/SQRT(D)	0.2*D	0.02	0.15	1.00	0.0	0.07	0.99	0.0	0.06	0.67	0.0	0.0	0.47
BETA	0	1.0*D	0.05	0.16	1.00	0.05	0.15	0.99	0.06	0.10	0.85	0.07	0.07	0.60
	1/SQRT(D)	1.0*D	0.04	0.18	1.00	0.03	0.10	0.99	0.04	0.07	0.78	0.03	0.02	0.50
	2/SQRT(D)	1.0*D	0.02	0.15	1.00	0.0	0.08	0.98	0.01	0.04	0.67	0.0	0.01	0.45
UNIFORM	0	0.2*D	0.06	0.21	1.00	0.03	0.20	1.00	0.03	0.08	0.91	0.01	0.03	0.72
	1/SQRT(D)	0.2*D	0.04	0.23	1.00	0.02	0.15	1.00	0.05	0.08	0.84	0.01	0.02	0.61
	2/SQRT(D)	0.2*D	0.03	0.16	1.00	0.01	0.10	0.99	0.01	0.03	0.73	0.0	0.01	0.43
UNIFORM	0	1.0*D	0.03	0.15	1.00	0.04	0.14	0.99	0.05	0.11	0.86	0.05	0.08	0.60
	1/SQRT(D)	1.0*D	0.05	0.19	1.00	0.05	0.12	1.00	0.04	0.09	0.79	0.07	0.07	0.54
	2/SQRT(D)	1.0*D	0.02	0.12	1.00	0.0	0.10	0.99	0.03	0.07	0.65	0.0	0.04	0.45
EXPONENTIAL	0	0.2*D	0.06	0.20	1.00	0.04	0.19	1.00	0.03	0.11	0.90	0.01	0.03	0.76
	1/SQRT(D)	0.2*D	0.05	0.22	1.00	0.02	0.12	0.99	0.02	0.09	0.83	0.0	0.01	0.60
	2/SQRT(D)	0.2*D	0.01	0.16	1.00	0.01	0.09	0.99	0.0	0.06	0.71	0.0	0.0	0.43
EXPONENTIAL	0	1.0*D	0.05	0.17	1.00	0.04	0.14	0.99	0.06	0.10	0.86	0.05	0.08	0.64
	1/SQRT(D)	1.0*D	0.03	0.20	1.00	0.02	0.15	0.99	0.03	0.09	0.83	0.02	0.05	0.64
	2/SQRT(D)	1.0*D	0.04	0.16	1.00	0.0	0.05	0.99	0.02	0.05	0.73	0.0	0.01	0.43
POWER USING NO PRIOR INFORMATION			0.05	0.15	1.00	0.04	0.12	0.98	0.07	0.09	0.64	0.04	0.05	0.21
POWER USING NULL PRIOR INFORMATION[2]			0.03	0.11	0.98	0.01	0.07	0.94	0.01	0.03	0.44	0.0	0.00	0.03

[1] THE EXPERIMENTER ADDS PRIOR KNOWLEDGE DISTRIBUTED N(0,D). POWER IS GIVEN FOR PRIORS DISTRIBUTED AS ABOVE.

[2] IN CALCULATING THE ESTIMATOR OF B, USING NULL PRIOR INFORMATION IS EQUIVALENT TO ADDING THE INVERSE OF D TO X'X.

TABLE 4.18. POWER OF THE 'T' TEST FOR H(0), B(1) = 0, AS IT RELATES TO CORRELATION AND PRIOR KNOWLEDGE (UNDERESTIMATION OF THE TRUE PRIOR VARIANCE)

D, THE PRIOR VARIANCE USED, IS 1000

DISTRIBUTION	CORRECT PRIORS[1] MEAN	VARIANCE	RHO = 0.00 B(1)=0,	1,	5	RHO = 0.50 B(1)=0,	1,	5	RHO = 0.85 B(1)=0,	1,	5	RHO = 0.96 B(1)=0,	1,	5
NORMAL	0	1.0*D	0.04	0.16	1.00	0.04	0.14	0.98	0.07	0.10	0.65	0.04	0.05	0.24
	1/SQRT(D)	1.0*D	0.05	0.11	1.00	0.01	0.09	0.98	0.03	0.08	0.52	0.01	0.02	0.16
	2/SQRT(D)	1.0*D	0.02	0.07	0.96	0.01	0.05	0.87	0.02	0.04	0.36	0.02	0.01	0.12
NORMAL	0	5.0*D	0.01	0.09	0.95	0.02	0.06	0.88	0.02	0.05	0.41	0.01	0.02	0.11
	1/SQRT(D)	5.0*D	0.0	0.05	0.92	0.0	0.01	0.86	0.01	0.04	0.33	0.0	0.01	0.04
	2/SQRT(D)	5.0*D	0.01	0.04	0.85	0.0	0.03	0.70	0.01	0.03	0.30	0.0	0.0	0.10
BETA	0	1.0*D	0.05	0.17	1.00	0.03	0.14	0.98	0.06	0.08	0.66	0.05	0.06	0.24
	1/SQRT(D)	1.0*D	0.02	0.12	1.00	0.02	0.08	0.98	0.07	0.08	0.48	0.02	0.02	0.15
	2/SQRT(D)	1.0*D	0.0	0.07	0.98	0.0	0.01	0.90	0.02	0.03	0.37	0.01	0.01	0.08
BETA	0	5.0*D	0.01	0.08	0.94	0.01	0.06	0.88	0.03	0.05	0.43	0.02	0.03	0.13
	1/SQRT(D)	5.0*D	0.02	0.05	0.91	0.0	0.06	0.88	0.01	0.04	0.42	0.0	0.0	0.10
	2/SQRT(D)	5.0*D	0.01	0.04	0.86	0.0	0.02	0.73	0.02	0.03	0.23	0.0	0.01	0.04
UNIFORM	0	1.0*D	0.04	0.15	1.00	0.03	0.12	0.98	0.06	0.10	0.63	0.04	0.05	0.24
	1/SQRT(D)	1.0*D	0.02	0.11	0.99	0.0	0.05	0.96	0.04	0.06	0.51	0.01	0.03	0.13
	2/SQRT(D)	1.0*D	0.01	0.06	0.96	0.0	0.02	0.88	0.01	0.06	0.37	0.01	0.0	0.08
UNIFORM	0	5.0*D	0.01	0.06	0.98	0.01	0.05	0.90	0.02	0.06	0.37	0.02	0.02	0.10
	1/SQRT(D)	5.0*D	0.0	0.04	0.91	0.0	0.03	0.80	0.02	0.04	0.27	0.01	0.01	0.05
	2/SQRT(D)	5.0*D	0.01	0.03	0.89	0.01	0.04	0.76	0.01	0.03	0.31	0.01	0.0	0.04
EXPONENTIAL	0	1.0*D	0.05	0.16	1.00	0.05	0.15	0.98	0.08	0.09	0.67	0.04	0.06	0.23
	1/SQRT(D)	1.0*D	0.05	0.15	0.99	0.02	0.13	0.97	0.07	0.08	0.58	0.01	0.02	0.13
	2/SQRT(D)	1.0*D	0.0	0.06	0.98	0.0	0.04	0.91	0.02	0.05	0.37	0.0	0.0	0.08
EXPONENTIAL	0	5.0*D	0.02	0.09	0.94	0.02	0.07	0.90	0.03	0.07	0.48	0.02	0.03	0.18
	1/SQRT(D)	5.0*D	0.03	0.10	0.91	0.03	0.06	0.85	0.05	0.07	0.37	0.03	0.02	0.11
	2/SQRT(D)	5.0*D	0.04	0.07	0.85	0.02	0.06	0.76	0.04	0.06	0.35	0.01	0.01	0.12
POWER USING NO PRIOR INFORMATION			0.05	0.15	1.00	0.04	0.12	0.98	0.07	0.09	0.64	0.04	0.05	0.21
POWER USING NULL PRIOR INFORMATION[2]			0.05	0.15	1.00	0.04	0.11	0.98	0.07	0.08	0.64	0.03	0.05	0.21

[1] THE EXPERIMENTER ADDS PRIOR KNOWLEDGE DISTRIBUTED N(0,D). POWER IS GIVEN FOR PRIORS DISTRIBUTED AS ABOVE.

[2] IN CALCULATING THE ESTIMATOR OF B, USING NULL PRIOR INFORMATION IS EQUIVALENT TO ADDING THE INVERSE OF D TO X'X.

TABLE 4.19. POWER OF THE 'T' TEST FOR H(0), B(1) = 0, AS IT RELATES TO CORRELATION AND PRIOR KNOWLEDGE (UNDERESTIMATION OF THE TRUE PRIOR VARIANCE)

D, THE PRIOR VARIANCE USED, IS 100

DISTRIBUTION	CORRECT PRIORS[1] MEAN	VARIANCE	RHO = 0.00 B(1)=0,	1,	5	RHO = 0.50 B(1)=0,	1,	5	RHO = 0.85 B(1)=0,	1,	5	RHO = 0.96 B(1)=0,	1,	5
NORMAL	0	1.0*D	0.06	0.17	1.00	0.04	0.14	0.98	0.06	0.08	0.69	0.06	0.07	0.29
	1/SQRT(D)	1.0*D	0.03	0.15	1.00	0.01	0.08	0.98	0.07	0.08	0.56	0.02	0.02	0.21
	2/SQRT(D)	1.0*D	0.02	0.06	0.99	0.0	0.03	0.90	0.0	0.04	0.38	0.0	0.0	0.07
NORMAL	0	5.0*D	0.02	0.06	0.96	0.01	0.05	0.90	0.04	0.04	0.48	0.04	0.04	0.23
	1/SQRT(D)	5.0*D	0.03	0.08	0.97	0.01	0.06	0.90	0.03	0.05	0.38	0.04	0.03	0.15
	2/SQRT(D)	5.0*D	0.03	0.06	0.87	0.01	0.04	0.71	0.03	0.04	0.28	0.02	0.02	0.15
BETA	0	1.0*D	0.05	0.16	1.00	0.04	0.12	0.98	0.07	0.08	0.68	0.06	0.06	0.28
	1/SQRT(D)	1.0*D	0.06	0.13	1.00	0.01	0.11	0.98	0.05	0.08	0.58	0.01	0.01	0.22
	2/SQRT(D)	1.0*D	0.01	0.07	0.97	0.0	0.05	0.92	0.01	0.06	0.41	0.0	0.01	0.10
BETA	0	5.0*D	0.02	0.07	0.96	0.01	0.06	0.90	0.02	0.06	0.45	0.05	0.05	0.19
	1/SQRT(D)	5.0*D	0.0	0.06	0.91	0.0	0.05	0.80	0.03	0.05	0.32	0.04	0.02	0.18
	2/SQRT(D)	5.0*D	0.0	0.07	0.84	0.0	0.03	0.75	0.03	0.03	0.29	0.03	0.04	0.12
UNIFORM	0	1.0*D	0.04	0.16	1.00	0.02	0.14	0.99	0.06	0.09	0.70	0.04	0.04	0.30
	1/SQRT(D)	1.0*D	0.02	0.14	1.00	0.01	0.14	0.97	0.07	0.07	0.53	0.04	0.03	0.21
	2/SQRT(D)	1.0*D	0.01	0.09	0.96	0.01	0.03	0.88	0.04	0.06	0.42	0.01	0.01	0.10
UNIFORM	0	5.0*D	0.01	0.06	0.96	0.01	0.05	0.90	0.03	0.05	0.42	0.05	0.07	0.20
	1/SQRT(D)	5.0*D	0.01	0.05	0.94	0.0	0.03	0.79	0.02	0.04	0.31	0.04	0.04	0.19
	2/SQRT(D)	5.0*D	0.01	0.03	0.91	0.0	0.04	0.77	0.03	0.01	0.36	0.0	0.0	0.13
EXPONENTIAL	0	1.0*D	0.06	0.17	1.00	0.04	0.14	0.98	0.08	0.10	0.70	0.05	0.06	0.29
	1/SQRT(D)	1.0*D	0.05	0.14	0.99	0.02	0.11	0.97	0.05	0.10	0.60	0.03	0.02	0.22
	2/SQRT(D)	1.0*D	0.01	0.09	0.97	0.01	0.04	0.91	0.01	0.05	0.37	0.01	0.0	0.13
EXPONENTIAL	0	5.0*D	0.02	0.08	0.92	0.02	0.07	0.88	0.03	0.05	0.50	0.05	0.04	0.20
	1/SQRT(D)	5.0*D	0.06	0.11	0.95	0.02	0.10	0.89	0.08	0.07	0.51	0.05	0.05	0.21
	2/SQRT(D)	5.0*D	0.01	0.08	0.88	0.0	0.06	0.82	0.04	0.03	0.42	0.01	0.02	0.15
POWER USING NO PRIOR INFORMATION			0.05	0.15	1.00	0.04	0.12	0.98	0.07	0.09	0.64	0.04	0.05	0.21
POWER USING NULL PRIOR INFORMATION[2]			0.04	0.14	1.00	0.04	0.10	0.98	0.05	0.07	0.59	0.02	0.02	0.16

[1] THE EXPERIMENTER ADDS PRIOR KNOWLEDGE DISTRIBUTED N(O,D). POWER IS GIVEN FOR PRIORS DISTRIBUTED AS ABOVE.

[2] IN CALCULATING THE ESTIMATOR OF B, USING NULL PRIOR INFORMATION IS EQUIVALENT TO ADDING THE INVERSE OF D TO X'X.

TABLE 4.20. POWER OF THE 'T' TEST FOR H(0), B(1) = 0, AS IT RELATES TO CORRELATION AND PRIOR KNOWLEDGE (UNDERESTIMATION OF THE TRUE PRIOR VARIANCE)

D, THE PRIOR VARIANCE USED, IS 10

DISTRIBUTION	CORRECT PRIORS[1] MEAN	VARIANCE	RHO = 0.00 B(1)=0,	1,	5	RHO = 0.50 B(1)=0,	1,	5	RHO = 0.85 B(1)=0,	1,	5	RHO = 0.96 B(1)=0,	1,	5
NORMAL	0	1.0*D	0.04	0.17	1.00	0.02	0.16	1.00	0.05	0.09	0.87	0.04	0.06	0.62
	1/SQRT(D)	1.0*D	0.04	0.23	1.00	0.03	0.11	1.00	0.05	0.05	0.76	0.04	0.06	0.49
	2/SQRT(D)	1.0*D	0.03	0.17	1.00	0.02	0.06	0.98	0.01	0.04	0.60	0.0	0.01	0.31
NORMAL	0	5.0*D	0.04	0.10	0.97	0.03	0.11	0.89	0.11	0.11	0.64	0.18	0.19	0.50
	1/SQRT(D)	5.0*D	0.02	0.11	0.97	0.03	0.11	0.91	0.07	0.15	0.71	0.10	0.17	0.56
	2/SQRT(D)	5.0*D	0.02	0.14	0.93	0.04	0.10	0.82	0.08	0.13	0.57	0.13	0.19	0.48
BETA	0	1.0*D	0.05	0.16	1.00	0.05	0.15	0.99	0.06	0.10	0.85	0.07	0.07	0.60
	1/SQRT(D)	1.0*D	0.04	0.18	1.00	0.03	0.10	0.99	0.04	0.07	0.78	0.03	0.02	0.50
	2/SQRT(D)	1.0*D	0.02	0.15	1.00	0.0	0.08	0.98	0.01	0.04	0.67	0.0	0.01	0.45
BETA	0	5.0*D	0.03	0.10	0.96	0.04	0.11	0.94	0.12	0.13	0.67	0.13	0.20	0.52
	1/SQRT(D)	5.0*D	0.02	0.12	0.99	0.01	0.12	0.95	0.08	0.10	0.61	0.11	0.12	0.44
	2/SQRT(D)	5.0*D	0.02	0.10	0.99	0.01	0.08	0.92	0.04	0.05	0.58	0.09	0.09	0.38
UNIFORM	0	1.0*D	0.03	0.15	1.00	0.04	0.14	0.99	0.05	0.11	0.86	0.06	0.08	0.60
	1/SQRT(D)	1.0*D	0.05	0.19	1.00	0.05	0.12	1.00	0.04	0.09	0.79	0.07	0.07	0.54
	2/SQRT(D)	1.0*D	0.02	0.12	1.00	0.0	0.10	0.99	0.03	0.07	0.65	0.0	0.04	0.45
UNIFORM	0	5.0*D	0.03	0.09	0.96	0.04	0.11	0.90	0.13	0.14	0.63	0.29	0.29	0.48
	1/SQRT(D)	5.0*D	0.04	0.08	0.99	0.04	0.07	0.95	0.10	0.13	0.54	0.22	0.20	0.42
	2/SQRT(D)	5.0*D	0.03	0.11	0.98	0.03	0.10	0.85	0.07	0.14	0.57	0.14	0.16	0.48
EXPONENTIAL	0	1.0*D	0.05	0.17	1.00	0.04	0.14	0.99	0.06	0.10	0.86	0.05	0.08	0.64
	1/SQRT(D)	1.0*D	0.03	0.20	1.00	0.02	0.15	0.99	0.03	0.09	0.83	0.02	0.05	0.64
	2/SQRT(D)	1.0*D	0.04	0.16	1.00	0.0	0.05	0.99	0.02	0.05	0.73	0.0	0.01	0.43
EXPONENTIAL	0	5.0*D	0.03	0.07	0.97	0.03	0.11	0.93	0.08	0.13	0.69	0.17	0.18	0.50
	1/SQRT(D)	5.0*D	0.06	0.11	0.97	0.03	0.10	0.94	0.04	0.05	0.70	0.11	0.11	0.54
	2/SQRT(D)	5.0*D	0.02	0.13	0.93	0.01	0.13	0.86	0.04	0.07	0.61	0.04	0.06	0.48
POWER USING NO PRIOR INFORMATION			0.05	0.15	1.00	0.04	0.12	0.98	0.07	0.09	0.64	0.04	0.05	0.21
POWER USING NULL PRIOR INFORMATION[2]			0.03	0.11	0.98	0.01	0.07	0.94	0.01	0.03	0.44	0.0	0.00	0.03

[1] THE EXPERIMENTER ADDS PRIOR KNOWLEDGE DISTRIBUTED N(0,D). POWER IS GIVEN FOR PRIORS DISTRIBUTED AS ABOVE.

[2] IN CALCULATING THE ESTIMATOR OF B, USING NULL PRIOR INFORMATION IS EQUIVALENT TO ADDING THE INVERSE OF D TO X'X.

4.2.1. The Effect of Adding Correct Prior Knowledge

Results displayed in Tables 4.15 through 4.20 indicate that if prior variance is as low as ten times the residual variance, then when the independent variables are highly correlated, there is a significant power gain, for large non-centrality parameters. In Table 4.17 when the β_1 associated with the non-centrality parameter is 5, if

1. $\rho = 0.85$, the "no prior" power is 0.64, and the power using correct prior knowledge with variance 10 is 0.87.

2. $\rho = 0.96$, the "no prior" power is 0.21, and the power using correct prior knowledge with variance 10 is 0.62.

Size is not changed, of course, and no significant power changes are noted when the prior variance is larger than 10.

It should be noted that in Section 4.1 the effect of prior knowledge on mean square error of $\hat{\beta}_1$ was discussed not considering how this effect was tempered by the estimation of σ^2. For power calculations, of course, both σ^2 and $\underline{\beta}$ were estimated.

4.2.2. Distribution Specification Errors When Prior Knowledge Is Correct in Its Mean and Variance

Statistics were generated with prior knowledge distributed from three non-normal distributions. The beta and uniform distributions used were symmetric. It is of interest to note that when the mean and variance specifications are correct, no significant power modifications are evident, even if the distribution is unsymmetric, as in the exponential case.

4.2.3. The Effect of Errors in Specification of Prior Variances and Means When the Prior Information Is Normally Distributed

Tables 4.15 through 4.22 are displayed in this section to allow a determination of

1. Effects when variance is correctly specified and means are incorrect.

2. Effects when the mean is correctly specified and variances are incorrect.

3. The effect of errors in specifying both means and variances of the priors.

In Tables 4.15 - 4.20, the assumed prior variance, $D = dI$, is perturbed by a five factor in both directions to allow consideration of effects of over and under estimation of prior variances. Since there was no evidence of distributional error effects, and since more results where $d = d_0$ were desired, Tables 4.21 and 4.22 were constructed as follows:

1. Results were averaged over all distributions for the cases $d = 10$, 100, and 1000 in Tables 4.15 through 4.20, and the elements of Table 4.21 were obtained by interpolation. These elements, and those of Table 4.22, are therefore based on 1000 simulations when $md^{-1/2} = 0$, and on 400 simulations when biases are non-zero.

2. Table 4.22 was extracted from Tables 4.15 through 4.21 to show more clearly the effect of over and under estimation, based on the true prior variances. Whereas Tables 4.15 through 4.20 show perturbation of the assumed prior variances, Table 4.22 uses the true prior variances as a base.

When overestimation of prior variance has been the error situation, and the mean is specified correctly, Table 4.22 shows that at the level $d_0 = 20$, a decrease in the power ordinate at $\beta_1 = 5$ is evident (.48 to .32) when correlation is .96. When $d_0 = 200$, no change is noted. Tables 4.15 through 4.17 expand the view of the power curve. For the normally distributed sections, no changes are noted for d = 1000 or d = 100, but when d = 10 ($d_0 = 2$) a slight decrease in the ordinate at $\beta_1 = 5$ can be seen. The size of the test ($\beta_1 = 0$) remains stable.

In the case of underestimation, with correct mean specification, Table 4.22 shows, at $\beta_1 = 5$, when $d_0 = 50$, an increase in the ordinate for high correlations, while no changes are noted when correlation is not greater than 0.50. At $d_0 = 500$, no significant differences can be noted.

If variance is correctly specified and the mean of the prior is incorrect, one can see, in Tables 4.15 through 4.22, the following:

1. Table 4.21: No significant changes are noted for correlations below 0.85; when $\rho = 0.96$, and the bias is twice the standard deviation of the prior used, a large decrease in the power ordinate is noted at $\beta_1 = 5$.

2. Tables 4.15 through 4.17: The result of mean specification errors seems to imply a general lowering of the whole power curve, and this tendency increases with correlation.

Finally, when both types of errors are in evidence, one sees, in Table 4.22, that when $d_0 = 20$, and the experimenter overestimates, the

Table 4.21. Ordinates of the power curve when $\beta_1 = 5$

$md^{-1/2}$	Correlation 0.00	0.50	0.85	0.96	$d=d_0$
No Prior---	1.00	0.98	0.64	0.21	
0	1.00	0.99	0.78	0.48	
1	1.00	0.99	0.70	0.40	20
2	1.00	0.95	0.54	0.27	
0	1.00	0.99	0.71	0.34	
1	1.00	0.98	0.60	0.27	50
2	0.99	0.91	0.44	0.14	
0	1.00	0.99	0.68	0.28	
1	1.00	0.98	0.56	0.20	200
2	0.99	0.90	0.39	0.10	
0	1.00	0.98	0.66	0.26	
1	1.00	0.97	0.52	0.17	500
2	0.97	0.89	0.37	0.09	

percentage decrease in the power ordinate becomes larger as $md^{-1/2}$ increases, while this is not evident when $d_0 = 200$. When the experimenter underestimates, the same can be said about the percentage increase; it becomes larger as $md^{-1/2}$ increases, at the level $d_0 = 50$, but no change can be noted when $d_0 = 500$; bias reaches the order of twice the standard deviation of the prior before the ordinate becomes smaller than the no prior power ordinate.

It should again be noted that the results in Tables 4.15 through 4.22 are ordinate results; sizes of the test are changing, but although they are lower when $md^{-1/2}$ is two, as seen in Tables 4.15 through 4.20, the change may not be significantly large.

Table 4.22. The ordinates of the power curve at $\beta_1 = 5$ when variance and mean specifications are incorrect[1]

$md^{-1/2}$	Correlation 0.00	0.50	0.85	0.96	d	d_0
No Prior------	1.00	0.98	0.64	0.21		
			Overestimation			
0	1.00	0.99	0.77	0.32		
1	0.99	0.98	0.59	0.22	100	20
2	0.99	0.95	0.41	0.07		
0	1.00	0.99	0.78	0.48		
1	1.00	0.99	0.70	0.40	20	20
2	1.00	0.95	0.54	0.27		
0	1.00	0.99	0.74	0.29		
1	1.00	0.98	0.56	0.17	1000	200
2	0.99	0.93	0.38	0.06		
0	1.00	0.99	0.68	0.28		
1	1.00	0.98	0.56	0.20	200	200
2	0.99	0.90	0.39	0.10		
			Underestimation			
0	0.97	0.92	0.66	0.50		
1	0.97	0.94	0.64	0.49	10	50
2	0.96	0.86	0.58	0.46		
0	1.00	0.99	0.71	0.34		
1	1.00	0.98	0.60	0.27	50	50
2	0.99	0.91	0.44	0.14		
0	0.95	0.90	0.47	0.21		
1	0.94	0.85	0.38	0.18	100	500
2	0.87	0.76	0.33	0.14		
0	1.00	0.98	0.66	0.26		
1	1.00	0.97	0.52	0.17	500	500
2	0.97	0.89	0.37	0.09		

4.2.4. Effect in the Null Prior Case

No change in the power ordinate at any level of the non-centrality parameter can be noted until prior variance is on the order of 10. At 10 (Table 4.17) both size and the power ordinates are smaller than the use of $\hat{\underline{\beta}}$ will provide, for correlations of 0.85 and 0.96.

A further simulation study was run with two levels of β_2, the non-hypothesized parameter, so that its effect on the test resulting from the use on null prior knowledge could be studied. Table 4.23 was constructed using fifty simulations, but it should be noted that the same y's--*i.e.*, observations--were used in calculating both statistics, *i.e.*, the no prior $\hat{\underline{\beta}}$ and the null prior β_h.

As can be seen, the parameter β_2 does change power ordinates; size and power ordinates are lower than those from the no prior test. Note where $\beta_2 = 5$ and the prior standard deviation is 10; a change caused by the level of β_2 should be evident from our earlier considerations, since Hoerl's estimator is biased by $-\underline{\beta}$.

Table 4.23. Null prior knowledge, the power of the test for $H_0: \beta_1=0$, and the effect of the level of β_2; the prior variance used is d = 10, and the power of the least squares statistic (no prior) is given in parentheses

(β_1,β_2)	Correlation 0.00		0.50		0.85		0.96	
0,0	0.02	(0.10)	0.02	(0.04)	0.00	(0.10)	0.00	(0.04)
0,5	0.00		0.00		0.02		0.00	
1,0	0.08	(0.18)	0.06	(0.12)	0.02	(0.12)	0.00	(0.02)
1,5	0.06		0.04		0.06		0.02	
5,0	0.98	(1.00)	0.90	(1.00)	0.30	(0.60)	0.02	(0.16)
5,5	0.98		0.82		0.36		0.08	

5. SUMMARY AND CONCLUSIONS

5.1. General Remarks and Conclusions

In ill-conditioned situations, which occur often in practice, an experimenter is faced with inversion routines which give incorrect inverses, and with parameter estimates which, although unbiased, have such high variance that lack of bias is unimportant. Often few simple options are available to him. He is surrounded by the many other problems associated with designing a model in an industrial atmosphere. Time and available resources play as important a role, often, as do correct theoretical procedures. This is an unfortunate, but seemingly lasting, state of affairs.

The experimenter's options may be to

1. Redesign the experiment so as to assure near orthogonality of independent variables,

2. Use more sophisticated computer codes, applying high precision in calculations to numbers which are not accurate in themselves,

3. Eliminate several variables which seem important,

4. Delimit the range of interest,

5. Examine non-linear models.

An added possibility is to precondition the model in the manner of Hoerl, hoping that bias will be an unimportant factor. Bias will be important, however, if the magnitude of the parameter set, $\underline{\beta}$, is large. In this case, the only test that can be applied is the one which is always most crucial, that of experience in using the resultant model. Experience with Hoerl's conditioning technique shows that the resultant models serve much better than do least squares models when the condition of the information matrix is poor.

If prior knowledge is available, then the same preconditioning of the model is automatic. Theoretical foundations are also at hand. In the following, conclusions concerning the estimation of $\underline{\beta}$, the estimation of σ^2, prediction, and the use of prior knowledge on σ^2 are given.

From the results discussed in Section 4.1.1, it is evident that properties of estimators of $\underline{\beta}$ are quite robust to the addition of incorrect prior knowledge, with certain restrictions.

1. If the true bias of the prior mean specification is small compared to the standard deviation of the prior, the experimenter will not make errors in judgment.
2. If the correct variance of the prior is smaller than the actual variance applied, errors will be conservative in nature.

With these restrictions, the addition of prior knowledge, even if incorrect, is of benefit to the experimenter. Mean square error of the estimator of $\underline{\beta}$ is smaller than the variance of the "no prior" estimator, and the variance of the estimator as that variance is seen by the experimenter is larger than the true mean square error. The robustness of properties is dependent on correlation. If X'X is well-conditioned, prior knowledge has much less effect, both beneficial and non-beneficial, than is the case when X'X is nearly singular. Correctness of prior specification is more important when X'X is well-conditioned.

Prior information about $\underline{\beta}$ is also important in estimating σ^2. Properties of such estimates are not related to the size of the prior variance, or to correlation. As opposed to the situation when estimating $\underline{\beta}$, the impact of applying incorrect prior knowledge on the properties of the estimator of σ^2 does not necessarily become greater with correlation. This would imply that the experimenter might well achieve a better estimator for $\underline{\beta}$ at the expense of estimating σ^2 poorly. When the prior concerning $\underline{\beta}$ has no bias, although overestimation of its variance is to be preferred, rather than underestimation, for estimation of σ^2, the bad effects of underestimation decrease as the prior variance used decreases. That was seen to be true both when prior knowledge on $\underline{\beta}$ was used, and when prior knowledge on $\underline{\beta}$ and σ^2 was added. The opposite is the situation when $\underline{\beta}$ estimates are considered. Thus, a balance between the estimation of $\underline{\beta}$ and σ^2 is possible by the correct choice of prior variance, when the true prior variance is unknown. This is to be investigated more fully in the program outlined in Section 5.3.

Again, however, if biases in the mean of the prior are small, and if the experimenter overestimates prior variance, his σ^2 estimator will be better than the "no prior" estimator.

The same evidences as given above apply to the variance of the prediction. Predictive properties are more robust and less affected by the addition of incorrect prior knowledge when X'X is well-conditioned, as was the case for $\underline{\beta}$ estimation. When X'X is ill-conditioned, the benefits are more extreme, and overestimation of prior variance is an error that the experimenter can make with some impunity.

Adding prior knowledge on σ^2 requires more investigation. However, of the error types which can be made, errors of underestimation of the variance on $\underline{\beta}$ prior knowledge seem most important, and high biases in the mean of that prior are next in importance. The bad effects of these errors can be overcome by the addition of highly valid information on σ^2. Prior specification itself is of importance, especially if the prior is quite different from χ^2, as is the exponential.

For hypothesis testing on $\underline{\beta}$, the power simulations showed that benefits are always gained, i.e., the power ordinates are increased, if the prior knowledge is correct. All effects noted increase with correlation. No noticeable effect can be seen when prior knowledge is not normally distributed. When the experimenter overestimates, power ordinates decrease at lower levels of prior variance (less than $50\sigma^2$). The decrease is larger when means are biased more than σ. When he underestimates, power ordinates are increased at low levels of prior variance, and the increase is larger when mean biases are large. Correct variance specification with large mean biases imply a decrease in the power ordinates along the whole power curve. The use of a null prior results in a general reduction of the power curve at all points.

It is felt by the writer that when ill-condition is present, the robustness of estimator properties to the addition of incorrect prior knowledge is evident, when certain safeguards are taken. The experimenter should attach higher variances to his prior knowledge than he believes are correct. Then biases will be ineffective unless they are so large as to be near the standard deviation of the prior added, in size.

Better experimental design would be the correct answer to the experimenter's problems, but for many years, considering data gathering techniques used in industry, the need to build models which describe processes, given highly correlated data, will seem necessary to engineers. Properly acquired prior knowledge, and correct application of this knowledge to linear models, will aid experimenters.

5.2. An Analogy in Terms of Weights

When an experimenter is faced with analyzing data pertaining to models with highly correlated controllable variables, information available to him about the parameters of that model can be extremely useful. Proper usage of this information can result in better estimates of the parameters, even if the information is incorrect. As would be expected, the less faith the experimenter has in the validity of his prior knowledge, the lower weight he should attach to that knowledge. The results of this thesis show that, when $X'X$ is ill-conditioned, using a lower weight than the experimenter feels is correct will guard against errors. If the information used is biased, then unless the bias is as large as $1/\sqrt{w}$, where w is the weight attached to the information, then errors introduced will not be large. In this thesis, each element of the current experimental yields, $\underline{y}$, was given a weight of one, and weights were much smaller than 1. As w approaches one, the effect of the prior information increases, and incorrectness of this prior information is more important, in the production of serious errors. The addition of weights will have more effect on $\underline{\beta}$ estimates when X'X is badly conditioned than when it is well-conditioned.

5.3. Extensions Currently under Development

The questions considered in this thesis are: If the experimenter adds prior knowledge to the experimental yields at hand, what benefits are gained, and are these benefits the same when X'X is diagonal as when it is nearly singular? What happens if the prior knowledge is incorrect?

Both of the computer programs associated with these questions were based on the assumption of a mean, variance, and distributional form for the prior, and calculations were performed considering that these were both correct and incorrect. The two programs covered a wide range of situations. Eighty hours of time on an IBM 360 Model 40 was required. The twenty-four cases investigated in the power simulations (six models, and four correlations, each for three means, three correct prior variances, three assumed prior variances, and four correct distributional forms of prior knowledge) required 4.32 minutes of computing time per iteration, and 250 iterations were computed, excepting for the mean changes, where 100 iterations were computed.

With the results of this thesis in mind, the following investigation is under development. It is to be carried out in two phases. First, the investigation will generate X matrices as was done in this thesis. Then,

1. It will be assumed that prior knowledge has mean $\underline{\beta} + \underline{m}$, and variance $D_0\sigma^2$. The range of D_0 will be from 1 to 1000. $\underline{m}$ will range from $\underline{0}$ to $2D_0^{1/2}$.

2. The experimenter will apply knowledge with an assumed mean of $\underline{0}$, and an assumed variance from $0.1D_0$ to $10D_0$.

3. Only normal distributions will be considered.

4. Power will be determined by an approximation, in the manner developed by Patniak (1949). This will eliminate the computation time required by simulations.

5. Models with three parameters of varying importance will be considered, with correlations varying from 0 to .96.

6. Prior knowledge on σ^2 will not be considered.

7. n, the number of observations, will be included in the parameter set under observation.

The above deliminations of the problem considered in this thesis will allow a detailed scrutiny of the conclusions given here, in the ranges which now seem important.

Second, X matrices will be taken from field data, and the above steps will be followed. It is hoped that this will produce a useful computer code for the ill-conditioned situation which will have a sound theoretical basis; such a code would accept prior knowledge with statements pertaining to its validity. Least squares estimates and estimates using the prior knowledge would be produced, and the experimenter would be presented with ranges of possible errors in the estimates which would depend on the validity he had attached to the prior knowledge.

6. LIST OF REFERENCES

Abramowitz, M., and I. A. Stegun. 1965. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications, Inc., New York.

Anderson, R. L., and T. A. Bancroft. 1952. Statistical Theory in Research. McGraw-Hill Book Company, Inc., New York.

Box, G. E. P., and M. E. Muller. 1958. A note on the generation of normal deviates. Ann. Math. Stat. 29:610-611.

Chipman, J. S. 1964. On least squares with insufficient observations. Journal of the American Statistical Association 59: 1078-1111.

Cochran, W. G. 1934. The distribution of quadratic forms in a normal system, with applications to the analysis of variance. Proceedings of the Cambridge Philosophical Society 30:178.

Draper, N. 1963. Ridge analysis of response surfaces. Technometrics 5: 469-480.

Durbin, J. 1953. A note on regression when there is extraneous information about one of the coefficients. Journal of the American Statistical Association 48: 799-808.

Dykstra, O. 1966. The orthogonalization of undesigned experiments. Technometrics 8: 279-290.

Graybill, F. A. 1961. An Introduction to Linear Statistical Models. McGraw-Hill Book Company, Inc., New York.

Harvey, J. R., 1965. Fractional moments of a quadratic form in non-central normal random variates. Unpublished Ph.D. thesis, Department of Experimental Statistics, North Carolina State University at Raleigh. University Microfilm, Ann Arbor, Michigan.

Hoerl, A. E. 1959. Optimum solution of many variables equations. Chemical Engineering Progress 55: 11-32.

Hoerl, A. E. 1962. The application of ridge analysis to regression problems. Chemical Engineering Progress 58: 54-59.

Hutchinson, D. W. 1966. A new uniform pseudorandom number generator. Communications of the Association for Computing Machinery 9: 432-433.

I. B. M. Corporation. 1959. Random Number Generation and Testing. Reference Manual C20-8011, New York.

Imhof, J. P. 1961. Computing the distribution of quadratic forms in normal variables. Biometrika 48: 419-426.

Kendall, M. G., and A. Stuart. 1958. The Advanced Theory of Statistics, Vol. II. Charles Griffin and Company, London.

Lehmer, D. L. 1951. Mathematical methods in large-scale computing units. Ann. Comp. Lab. Harvard U. 26: 141-146.

MacLaren, M. D., and G. Marsaglia. 1965. Uniform random number generators. Journal of the Association for Computing Machinery 12: 83-89.

Patniak, P. B. 1949. The non-central χ^2 and F distributions and their applications. Biometrika 36: 202-232.

Stuart, A. 1958. Note: iterative solutions of likelihood equations. Biometrics 14: 128-130.

Wells, W. T., and R. L. Anderson. 1964. A discussion of best estimate of trajectory. Technical Staff Technical Memorandum Number 21, PAA5-35A, Air Force Missile Test Center, Air Force Systems Command, United States Air Force, Patrick Air Force Base, Florida.

7. APPENDICES

7.1. Prior Information

7.1.1. Selection of Prior Information When Regressions on the Same Model Are Available

If r independent regressions have been run on the same model, each having been based on n_i, i = 1,...,r, observations, then the prior knowledge ($\underline{\beta}_0$, c, and k) can be selected as follows:

For the ith regression,

$$\underline{y}_i = X_i\underline{\beta} + \underline{\varepsilon}_i, \ \underline{\varepsilon}_i \sim MN(\underline{0};\ \sigma^2 I_n),\ i = 1,\ldots,r.$$

Then $\quad \underline{\beta}_i = (X_i'X_i)^{-1}X_i'\underline{y}_i$.

Let $\quad (X_i'X_i)^{-1} = (c_{kj,k})$;

and let $\quad \beta_{0j} = \sum_{i=1}^{r} \hat{\beta}_{ij}/r$.

Since $\quad V\hat{\beta}_{ij} = c_{jj,i}\sigma^2$, for j = 1,...,p,

it follows that

$$V\beta_{0j} = \sigma^2 \sum_{i=1}^{r} c_{jj,i}/r^2 .$$

If the prior knowledge used by the experimenter is to be

$$\underline{\beta}_0 = \underline{\beta} + \underline{\delta},\ \underline{\delta} \sim MN(\underline{0};\ \sigma^2 D) ,$$

then D will be chosen to be

$$D = (\sum_{i=1}^{r} c_{jj,i}/r^2) = (d_j),\ j = 1,\ldots,p .$$

This method of choosing D ignores the correlation between the elements of $\underline{\hat{\beta}}$; one is saying that the elements of $\underline{\beta}_0$ are independent, and is determining $\underline{\beta}_0$ as the average of previous estimates of $\underline{\beta}$, $\underline{\hat{\beta}}_i$, where the elements of $\underline{\hat{\beta}}_i$ are not independent. Independent observations in the vector $(\underline{y}'; \underline{\beta}_0')$ are of aid in proving asymptotic normality of maximum likelihood estimators, but the lack of independence in $\underline{\beta}_0$ means that new estimates do not take full advantage of prior information. One could, since

$$\underline{\hat{\beta}}_i \sim MN[\underline{\beta};(X_i'X_i)^{-1}\sigma^2],$$

using standard assumptions, let $\underline{\beta}_0 = \sum_{i=1}^{r} \underline{\hat{\beta}}_i/r$ have variance

$$V\underline{\beta}_0 = \sigma^2 \sum_{i=1}^{r} (X_i'X_i)^{-1}/r^2 ,$$

since $\underline{\hat{\beta}}_i$ is independent of $\underline{\hat{\beta}}_{i'}$, when $i \neq i'$. In this thesis, elements of $\underline{\beta}_0$ are considered to be independent, a valuable consideration in the case of non-normal prior densities. For example, beta priors give rise to likelihood equations which cannot be solved by direct methods, so that asymptotic normality of the estimators is important as a criterion. For the problems considered here, iterative solutions of likelihood equations are not required; independence is desired because it provides diagonal matrices for augmentation of X'X, and this type of augmentation can be more directly related to conditioning of ill-conditioned matrices than can augmentation by a general positive definite matrix.

For the choice of c and k, when prior knowledge on σ^2 is used, the previously completed regressions provide

$$(n_i-p)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n_i-p} \quad .$$

Since the regressions are independent,

$$\sum_{i=1}^{r} (n_i-p)\hat{\sigma}_i^2/\sigma^2 \sim \chi^2_{\sum_i n_i-rp} \quad .$$

Thus, one would choose

$$c = \sum_{i=1}^{r} (n_i-p)\hat{\sigma}_i^2$$

and

$$k = \sum_{i=1}^{r} n_i - rp \ .$$

If the prior knowledge about σ^2 is

$$(c/\sigma^2 \sim \chi^2_k \quad ,$$

then the density is

$$f(c/\sigma^2) = \frac{(c/\sigma^2)^{\frac{k}{2}-1} e^{\frac{-c}{2\sigma^2}}}{2^{k/2}\Gamma(\frac{k}{2})} \quad , \quad c > 0 \ .$$

Transforming to a density on the variable c, one obtains equation (2.13);

$$h_1(c|\sigma^2) = \frac{c^{(k-2)/2} e^{-c/(2\sigma^2)}}{2^{k/2}\Gamma(\frac{k}{2})\ \sigma^k} \ ; \quad c > 0 \ .$$

For this density,

$$E(c) = k\sigma^2;$$

$$V(c) = 2k\sigma^4 \ .$$

7.1.2. The Beta Prior Density

A variate x, distributed as a beta variate in the interval $[c_1,c_2]$, is such that

$$f(x) = \frac{\Gamma(\theta_1+\theta_2)(x-c_1)^{\theta_1-1}(c_2-x)^{\theta_2-1}}{\Gamma(\theta_1)\Gamma(\theta_2)(c_2-c_1)^{\theta_1+\theta_2-1}}, \quad \theta_i > 0, \quad i = 1,2 \tag{7.1}$$

Then $Ex = \dfrac{c_2\theta_1 + c_1\theta_2}{\theta_1+\theta_2}$ and $Vx = \dfrac{(c_2-c_1)^2\theta_1\theta_2}{(\theta_1+\theta_2+1)(\theta_1+\theta_2)^2}$.

If $Ex = 0$, $c_2\theta_1 = -c_1\theta_2$; $Vx = d_0 = E^2(x)$ means that

$$\theta_2 = \frac{-c_2(c_1c_2+d_0)}{d_0(c_2-c_1)} \quad ; \quad \theta_1 = -c_1\theta_2/c_2 \; .$$

Since $c_1 < 0$ and $c_2 > 0$, $\theta_1 > 0$ and $\theta_2 > 0$ are achieved if $c_1c_2 < -d_0$. This, then, must hold for a valid beta density with mean zero. The simulations used $\theta_i \geq 1$. In general, beta variates will be taken from (7.1).

For simulations, symmetric beta densities are used. This means that

$$c_1 = -c_2, \text{ and}$$
$$\theta_1 = \theta_2 = (c_2^2 - d_0)/2d_0 \; .$$

For θ_1 to be positive, c_2 must be greater than $\sqrt{d_0}$, and for θ_1 to be greater than one, c_2 must be greater than $\sqrt{3d_0}$.

For the simulations used in the beta case, it was desired to obtain a deviate from a beta distribution close, in appearance, to the normal prior used. Since, in the normal case,

$$\underline{\delta} \sim MN(\underline{0};\ D_0\sigma^2),\ D_0 = d_0 I.$$

and σ^2 was unity, beta endpoints of $\pm\ 3\sigma$ were chosen to be $[-3\sqrt{d_0},\ 3\sqrt{d_0}]$, and $\theta_1 = \theta_2 = 4$. The maximum of this beta density is $.36\ d_0^{-1/2}$, as compared to the normal density maximum of $.16\ d_0^{-1}$.

7.1.3. <u>The Uniform Density in $[c_1, c_2]$</u>

In (7.1), $\theta_1 = \theta_2 = 1$.

$$f(x) = (c_2 - c_1)^{-1}.$$

If the variance of x is to be d_0, and the mean of x is zero,

$$c_2 = -c_1 = \sqrt{3d_0}\ .$$

That is, a rectangular distribution such that its mean is zero and its variance is d_0 is

$$f(x) = (2\sqrt{3d_0})^{-1},\ x \in [-\sqrt{3d_0},\ \sqrt{3d_0}]\ .$$

7.1.4. <u>The Exponential Prior Density</u>

The random variable x, distributed as an exponential variate, is such that

$$f(x) = \alpha e^{-\alpha x};\quad \alpha > 0,\ x > 0.$$

For the simulations, the desired density is

$$\delta \sim \exp(\text{mean zero, variance } d_0).$$

Therefore random variables will be generated from

$$f(x) = \frac{1}{\sqrt{d_0}} e^{-1-x/\sqrt{d_0}}, \quad d_0 > 0, \quad x > -\sqrt{d_0} .$$

7.1.5. The Prior Density for c/σ^2

When c/σ^2 is distributed exponentially with mean k, then

$$f(c/\sigma^2) = \frac{1}{k} e^{-c/(k\sigma^2)} \qquad k > 0, \quad c > 0 .$$

Transforming, one obtains

$$g(c|\sigma^2) = \frac{1}{\sigma^2 k} \quad e^{-c/(k\sigma^2)} \qquad .$$

Hence,

$$Ec = k\sigma^2, \text{ and}$$

$$Vc = k^2\sigma^4 .$$

7.2. Random Number Generation

Several pseudorandom number generators have been developed recently which are acceptable for use with computers. MacLaren and Marsaglia (1965) have put stringent tests on four general types of uniform random number generators. By their tests,

1. The mixed congruential method is unsatisfactory.
2. Multiplicative methods fail the test for triplets of uniform random numbers.
3. A combination of two congruential generators is satisfactory.
4. The use of a stored table of random numbers seems best.

The mixed congruential method uses the procedure

$$U_{i+1} = aU_i + c, \text{ modulo } M,$$

where M is 2^n for an n-bit binary machine,

a is near $\sqrt{M}$ (to give high periodicity),

and c is an odd integer.

If c = 0, the congruential method is termed multiplicative.

The combination of two congruential generators is a method which first fills a table with random deviates, using the mixed congruential method. If computing time is of concern, the high order bits of the random numbers in the table can be used to select the tabled deviates at random. Otherwise, a second mixed congruential generator is used to select from the table. Hutchinson (1966) notes that for the congruential methods, low-order bits of the generated deviate do not appear to be random, as are the high-order bits. If one requires random bits or digits, this is of consequence. Hutchinson tested a modification of the original multiplicative method given by Lehmer (1951) which solves this problem.

In the program used in this thesis, machine time and storage requirements implied that tables of random numbers could not be used, and that two generations of deviates would be unwise. Since the multiplicative method given by I. B. M. (1959) tested well according to the tables given by MacLaren and Marsaglia (1965), excepting for triplets, was very fast, and had a cycle of 2^{29} numbers, it was used for uniform deviates. Also, in this program random numbers were needed, and not random digits, so that the problem noted by Hutchinson (1966) did not arise.

Generators for all random numbers are given in Figures 7.1 through 7.5 below. Each generator is based on the generation of uniform deviates. Fortran was the programming aid used. Normal deviates were obtained by using an exact technique given by Box and Muller (1958). χ^2 deviates were obtained from summing squares of standard normal deviates. Beta deviates used the equation

$$\beta(m/2,n/2) = \chi_m^2/(\chi_m^2 + \chi_n^2) ,$$

where χ_m^2 is independent of χ_n^2. Exponential deviates invert the distribution function of the exponential density used in this thesis. That is,

$$y = \int_{-\infty}^{x} f(t)dt$$

is solved for x, where y is a uniform deviate; f(t) is given by (2.6).

```
      SUBROUTINE RANDU(IX, IY, YFL)
C        Computes uniformly distributed random numbers between zero
C        and one, YFL, and random integers between zero and two to the
C        thirty-first power, IY.  For the first entry IX should be any
C        odd integer with nine or less digits.  For succeeding entries,
C        IX should be the previous value of IY computed by this subrou-
C        tine.  Two to the twenty-ninth power terms are produced before
C        cycling occurs.  The Power Residue Method (I. B. M. Corporation,
C        1959) is used.
      IY = IX*65539
      IF(IY)5,6,6
    5 IY = IY + 2147483647 + 1
    6 YFL = IY
      YFL = YFL * .4656613E-09
      RETURN
```

Figure 7.1. A uniform random number generator

```
      SUBROUTINE RNOR(IX, N, E)
C        Generates N normal(0,1) random numbers.  On exit, E contains
C        these numbers.  Initially, IX should be set to an odd positive
C        integer.  IX is set for reentry by the subroutine.  N must be
C        even.  The Box exact normal technique is used (G. E. P. Box, and
C        M. E. Muller, 1958).  The Subroutine RANDU, above, is used.
      DIMENSION E(1)
      DOUBLE PRECISION E, W, U
      DO 1  I = 1,N
      CALL RANDU(IX,IY,RAN)
      E(I) = DBLE(RAN)
    1 IX = IY
      DO 2 I = 1,N,2
      IF(E(I))3,3,4
    4 U = DSQRT(-2.0D0*DLOG(E(I)))
    5 W = 6.283185307179586*E(I+1)
      E(I) = U*DCOS(W)
      E(I+1) = U*DSIN(W)
      GO TO 2
    3 U = 1.0D+12
      GO TO 5
    2 CONTINUE
      RETURN
```

Figure 7.2. A normal random number generator

```
      SUBROUTINE CHI(IX, N, CHI2)
C         Generates one chi-square deviate with N degrees of freedom.
C         The subroutine RNOR is used.  CHI2 is the result.
      DIMENSION E(1)
      DOUBLE PRECISION E, CHI2
      M = N-2*(N/2)
      IF(M)3,3,4
    4 M = N + 1
      GO TO 5
    3 M = N
    5 CALL RNOR(IX, M, E)
      CHI2 = 0.0D0
      DO 2 I = 1,N
    2 CHI2 = CHI2 + E(I)**2
      RETURN
```

Figure 7.3. A chi-square deviate generator

```
      SUBROUTINE BETA(IX, P, Q, B)
      DOUBLE PRECISION B, B1
C         Generates a beta(P,Q) variate, B.  P and Q must be multiples
C         of one-half.  The method used is
C
C                      2
C                     X
C                      2p
C                  ----------  = beta (p,q), where the two
C                    2     2
C                   X   + X
C                    2p    2q
C         chi-square variates are independent.  The Subroutine CHI is
C         used.  IX is set according to instructions given in the
C         Subroutine RANDU.
      A1 = .001
      M = IFIX(P + P + A1)
      N = IFIX(Q + Q + A1)
      CALL CHI(IX, M, B)
      CALL CHI(IX, N, B1)
      B = B/(B + B1)
      RETURN
```

Figure 7.4. A beta deviate generator

```
      SUBROUTINE EXP(IX, V, E)
      DOUBLE PRECISION Y, E, V
C         Generates an exponential deviate with zero mean and variance V.
C         The density of x is
C
C                        -1-x/√v
C                       e
C                       ----------
C                          √v
C
C         This routine uses a uniform(0,1) deviate, Y, and transforms
C         using
C                        x = P^-1(Y).
C         The subroutine RANDU is used, and IX is set as noted in that
C         subroutine.  The resulting deviate is E.
      DN1 = 1.0D0
      CALL RANDU(IX, IY, Y1)
      IX = IY
      Y = DBLE(Y1)
      IF(Y1 - 1.0)2,1,1
    1 E = 1.0D+12
      GO TO 3
    2 E = -DSQRT(V)*(DN1+DLOG(DN1-Y))
    3 RETURN
```

Figure 7.5. An exponential deviate generator

7.3. Theorems[10] on Normal Quadratic and Bilinear Forms

In the following

$\underline{y}_n \sim MN(\underline{\alpha}_1, V_1)$, Rank $(V_1) = n$

$\underline{x}_p \sim MN(\underline{\alpha}_2, V_2)$, Rank $(V_2) = p$

S_n is a symmetric nxn matrix, of rank n

M_{np} is an nxp matrix of rank $p \leq n$

M_{pn} is a pxn matrix of rank $p \leq n$

$\underline{x}$ is independent of $\underline{y}$.

Theorem 1. $E(\underline{y}'S_n\underline{y}) = \underline{\alpha}_1'S_n\underline{\alpha}_1 + Tr(V_1S_n)$ (S_n need not be symmetric)

$V(\underline{y}'S_n\underline{y}) = 4\underline{\alpha}_1'S_nV_1S_n\underline{\alpha}_1 + 2Tr(V_1S_n)^2$

Theorem 2. $E(\underline{y}'M_{np}\underline{x}) = \alpha_1'M_{np}\alpha_2$

Theorem 3. $Cov(\underline{y}'S_n\underline{y}, \underline{x}'M_{pn}\underline{y}) = 2Tr(V_1S_n\underline{\alpha}_1\underline{\alpha}_2'M_{pn})$

$Cov(\underline{x}'S_p\underline{x}, \underline{x}'M_{pn}\underline{y}) = 2Tr(V_2S_p\underline{\alpha}_2\underline{\alpha}_1'M_{np}')$

Theorem 4. $V(\underline{x}'M_{pn}\underline{y}) = (\alpha_2'M_{pn}\alpha_1)^2$

$$+ \sum_{i=1}^{p} \sum_{j=1}^{n} m_{ij}^2 v_{1_{jj}} v_{2_{ii}}$$

where

$v_{s_{rr}}$ is the rrth diagonal element of V_s, s=1,2;

m_{ij} is the i,jth element of M_{pn}.

Theorem 5. $Cov(\underline{x}'S_p\underline{x}, \underline{y}'S_n\underline{y}) = 0$.

---

[10]These theorems were given in, or developed from, course notes given by Dr. H. L. Lucas, in a special topics course on the general linear model, at North Carolina State University at Raleigh, N. C.

7.4. Estimation When the Assumed Prior Densities Are Not Normal in Form

7.4.1. Beta Prior Densities

Referring to equations (2.26) and (2.27), which are to be solved when the prior information is beta distributed, $Z(\underline{\omega})$ and $Z'(\underline{\omega})$ are developed as follows. They are given for $p = 2$, and the iterative method (2.27) is discussed specifically for this case.

Iterative solutions of equations (2.26) make comparisons of small sample properties of estimates to asymptotic properties desirable. Here, although $E(\underline{\omega}) = \underline{\omega}$, asymptotically, (Kendall and Stuart, 1961) the possibility that a globally unique solution of $Z(\underline{\omega}) = \underline{0}$ will not result from the iterative procedure used exists. According to Stuart (1958), the use of $E(\partial^2 \log L/\partial\underline{\omega}^2)$ in $Z'(\underline{\omega})$, rather than $\partial^2 \log L/\partial\underline{\omega}^2$, may hasten convergence and may reduce the possibility of non-convergence. This thesis develops the matrix Z' using $\partial^2 \log L/\partial\underline{\omega}^2$; that is using, in effect, the observations $\underline{y}^*$, rather than $E(\underline{y}^*)$, in $Z'(\underline{\omega})$. This results in non-symmetry for Z'.

The following definitions allow concise statements of the equations $Z(\underline{\omega})$, and of $Z'(\underline{\omega})$. Let $\underline{\omega}$ be defined as

$$\underline{\omega}' = (\beta_1;\beta_2;\sigma^2).$$

Define

$$\underline{t}' = (1;\beta_1;\beta_2;\beta_1^2;\beta_1\beta_2;\beta_2^2;\beta_1^2\beta_2;\beta_1\beta_2^2;\beta_1^3;\beta_2^3;\sigma^2;\beta_1\sigma^2;\beta_2\sigma^2);$$

$$\partial t/\partial\beta_1 = \underline{t}_{\beta_1}$$

where $\underline{t}'_{\beta_1} = (0;1;0;2\beta_1;\beta_2;0;2\beta_1\beta_2;\beta_2^2;3\beta_1^2;0;0;\sigma^2;0)$.

Let $\partial\underline{t}/\partial\beta_2 = \underline{t}_{\beta_2}$

where $\underline{t}'_{\beta_2} = (0;0;1;0;\beta_1;2\beta_2;\beta_1^2;2\beta_1\beta_2;0;3\beta_2^2;0;0;\sigma^2)$;

$\partial\underline{t}/\partial\sigma^2 = \underline{t}_{\sigma^2}$ where $\underline{t}'_{\sigma^2} = (0;0;0;0;0;0;0;0;0;0;1;\beta_1;\beta_2)$.

Define the following vectors:

$$\underline{b}_1 = \begin{bmatrix} k_1\underline{y}'\underline{x}_1 \\ -s_{11}k_1+2\beta_{01}\underline{y}'\underline{x}_1 \\ -s_{21}k_1 \\ -2\beta_{01}s_{11}-\underline{y}'\underline{x}_1 \\ -2\beta_{01}s_{21} \\ 0 \\ s_{21} \\ 0 \\ s_{11} \\ 0 \\ 2\beta_{01}(\theta_{11}-1) \\ -2(\theta_{11}-1) \\ 0 \end{bmatrix} \quad \underline{b}_2 = \begin{bmatrix} k_2\underline{y}'\underline{x}_2 \\ -s_{12}k_2 \\ -s_{22}k_2+2\beta_{02}\underline{y}'\underline{x}_2 \\ 0 \\ -2\beta_{02}s_{12} \\ -2\beta_{02}s_{22}-\underline{y}'\underline{x}_2 \\ 0 \\ s_{12} \\ 0 \\ s_{22} \\ 2\beta_{02}(\theta_{12}-1) \\ 0 \\ -2(\theta_{12}-1) \end{bmatrix} \quad \underline{b}_3 = \begin{bmatrix} \underline{y}'\underline{y} \\ 2\underline{y}'\underline{x}_1 \\ 2\underline{y}'\underline{x}_2 \\ s_{11} \\ 2s_{12} \\ s_{22} \\ 0 \\ 0 \\ 0 \\ 0 \\ -n \\ 0 \\ 0 \end{bmatrix}$$

In the above,

$k_1 = c_{2_1}^2 - \beta_{01}^2$;

$k_2 = c_{2_2}^2 - \beta_{02}^2$;

s_{ij} = i,jth element of X'X;

$\underline{x}_j$ = jth column of X;

θ_{1j} is the parameter of the symmetric beta density, defined in section 7.1.2; j implies the element of the β vector under consideration.

Using the above definitions,

$$Z(\underline{\omega}) = \begin{bmatrix} \underline{b}_1'\underline{t} \\ \underline{b}_2'\underline{t} \\ \underline{b}_3'\underline{t} \end{bmatrix} ; \qquad Z'(\underline{\omega}) = \begin{bmatrix} \underline{b}_1'\underline{t}_{\beta_1} & \underline{b}_1'\underline{t}_{\beta_2} & \underline{b}_1'\underline{t}_{\sigma^2} \\ \underline{b}_2'\underline{t}_{\beta_1} & \underline{b}_2'\underline{t}_{\beta_2} & \underline{b}_2'\underline{t}_{\sigma^2} \\ \underline{b}_3'\underline{t}_{\beta_1} & \underline{b}_3'\underline{t}_{\beta_2} & \underline{b}_3'\underline{t}_{\sigma^2} \end{bmatrix} .$$

7.4.2. Exponential Prior Densities

The techniques given in Section 7.4.1 are used when the prior knowledge is distributed exponentially. In this case,

$$\text{LogL} = k - (n/2)\log\sigma^2 - (\sigma^2/2)(\underline{y}-X\underline{\beta})'(\underline{y}-X\underline{\beta})$$

$$+ \sum_{j=1}^{p} (-\sqrt{d_{0j}} - \beta_{0j} + \beta_j)/\sqrt{d_{0j}} \; .$$

Again, let $\underline{\omega}' = (\underline{\beta}';\sigma^2)$. Then $Z(\underline{\omega}) = \partial\log L/\partial\underline{\omega} = \underline{0}$ is as follows:

$$Z(\underline{\omega}) = \begin{bmatrix} \underline{y}'\underline{x}_j - \underline{\beta}'\underline{s}_j + \sigma^2/\sqrt{d_{0j}} \\ -n\sigma^2 + (\underline{y}-X\underline{\beta})'(\underline{y}-X\underline{\beta}) \end{bmatrix} = \underline{0}$$

where $j = 1,2$. $\underline{x}_j$ is the jth column of X; $\underline{s}_j$ is the jth column of X'X; d_{0j} is the variance of the prior information. Z is comprised of three equations when p is 2.

For the solution of the set of equations Z, the iteration carried out, from a suitable starting point $\underline{\omega}_0$, is

$$\underline{\omega}_{i+1} = \underline{\omega}_i + [Z'(\underline{\omega}_i)]^{-1} Z(\underline{\omega}_i) \; ,$$

where

$$Z'(\underline{\omega}) = \partial^2\text{LogL}/\partial\underline{\omega}^2 = \begin{bmatrix} -s_{11} & -s_{21} & 1/\sqrt{d_{01}} \\ -s_{12} & -s_{22} & 1/\sqrt{d_{02}} \\ k_1 & k_2 & -n \end{bmatrix} \; ;$$

here, s_{ij} is the i,jth element of X'X, and

$$k_1 = -2\underline{y}'\underline{x}_1 + 2s_{11}\beta_1 + 2s_{12}\beta_2$$

$$k_2 = -2\underline{y}'\underline{x}_2 + 2s_{21}\beta_1 + 2s_{22}\beta_2 \; .$$